Psychometrika

CONTENTS

FURTHER STUDIES ON THE MATHEMATICAL THEORY OF INTERACTION OF INDIVIDUALS IN A SOCIAL GROUP - - - - - - - - - - - - 225
N. RASHEVSKY

USE OF THE TEST SCORING MACHINE AND THE GRAPHIC ITEM COUNTER FOR STATISTICAL WORK - - - - - - - - - - - - - - 233
BENJAMIN S. BLOOM AND ARDIE LUBIN

ON DETERMINING THE RELIABILITY AND SIGNIFICANCE OF A TETRACHORIC COEFFICIENT OF CORRELATION - - - - - - - - - - - 243
J. P. GUILFORD AND THOBURN C. LYONS

A FACTORIAL STUDY OF AUDITORY FUNCTION - - 251
J. E. KARLIN

TEST SCORES EXAMINED WITH THE LEXIS RATIO - 281
HAROLD A. EDGERTON AND KENNETH F. THOMSON

DERIVATION AND APPLICATION OF A UNIT SCORING SYSTEM FOR THE STRONG VOCATIONAL INTEREST BLANK FOR WOMEN - - - - - - 289
BERTHA P. HARPER AND JACK W. DUNLAP

A REVIEW OF "AN INQUIRY INTO THE PREDICTION OF SECONDARY-SCHOOL SUCCESS," BY W. G. EMMETT - - - - - - - - - - - - - - 297
MAX D. ENGELHART

INDEX - - - - - - - - - - - - - - - - - - - 299

VOLUME SEVEN DECEMBER 1942 NUMBER FOUR

PSYCHOMETRIKA—VOL. 7, NO. 4
DECEMBER, 1942

FURTHER STUDIES ON THE MATHEMATICAL THEORY OF INTERACTION OF INDIVIDUALS IN A SOCIAL GROUP

N. RASHEVSKY
THE UNIVERSITY OF CHICAGO

A type of interaction of two active groups is considered, in which the opposition of each group increases as the success of the other increases. Some possible applications of this situation are discussed.

In previous papers (**1, 2, 3, 4**) we have studied various cases of interactions of groups of individuals based on different assumptions as to their behavior. In this paper we shall consider a still different case, which presents some interest. We pre-suppose the knowledge of the previous papers by the reader. The notations and terminology are essentially the same.

Let us consider a case of interaction of two active groups, of such a nature that each group opposes the behavior of the other the more strongly the greater the success of that other group. This success may naturally be measured by the product of two factors: the ratio of the number of passive individuals which exhibit a given behavior to the total number of passives and the average intensity of that behavior. Let the average intensity of behavior A be denoted by w_A. Then the total success of class A will be expressed by $\alpha_1 w_A x/N'$ where α_1 is a coefficient; or putting

$$\alpha_1 w_A = \varepsilon , \qquad (1)$$

that success will be measured by $\varepsilon x/N'$. Similarly, if w_B denotes the average intensity of behavior B, then the success is given by $\alpha_2 w_B y/N'$ or, putting

$$\alpha_2 w_B = \varepsilon', \qquad (2)$$

by $\varepsilon' y/N'$.

How to measure the quantities w_A and w_B is another question. We shall just consider the whole problem *in abstracto* and then give some possible concrete illustrations.

In accordance with the foregoing assumptions and using the same procedure as before (**1**, **3**), we put

$$
\begin{aligned}
a_0 &= a^*{}_0(1 + \varepsilon' \frac{y}{N'}) , \\
c_0 &= c_0{}^*(1 + \varepsilon \frac{x}{N'}) .
\end{aligned}
\tag{3}
$$

The first expression in (3) shows that the effort of class A increases as the success of class B increases. The second expression shows a corresponding thing for class B.

We now have

$$
\frac{dx}{dt} = a^*{}_0(1 + \varepsilon' \frac{y}{N'}) x_0 + ax - c^*{}_0(1 + \varepsilon \frac{x}{N'}) y_0 - ay , \tag{4}
$$

or, because of

$$
x + y = N', \tag{5}
$$

after rearrangements

$$
\begin{aligned}
\frac{dx}{dt} = {} & (2a - a^*{}_0\varepsilon' \frac{x_0}{N'} - c^*{}_0\varepsilon \frac{y_0}{N'}) x \\
& + [(a^*{}_0\varepsilon' \frac{x_0}{N'} - a) N' + a^*{}_0 x_0 - c^*{}_0 y_0] .
\end{aligned}
\tag{6}
$$

The value of x tends asymptotically to

$$
x = \frac{a^*{}_0\varepsilon' x_0 - aN' + a^*{}_0 x_0 - c^*{}_0 y_0}{a^*{}_0\varepsilon' \frac{x_0}{N'} + c^*{}_0\varepsilon \frac{y_0}{N'} - 2a} , \tag{7}
$$

and is positive under similar assumptions as made before **(1, 3)**. A similar expression is obtained for y and therefore

$$
\frac{x}{y} = \frac{a^*{}_0\varepsilon' x_0 - aN' + a^*{}_0 x_0 - c^*{}_0 y_0}{c^*{}_0\varepsilon y_0 - aN' + c^*{}_0 y_0 - a^*{}_0 x_0} . \tag{8}
$$

Equation (8) may be written:

$$
\frac{x}{y} = \frac{(a^*{}_0\varepsilon' + a^*{}_0) \frac{x_0}{y_0} - (c^*{}_0 + \frac{aN'}{y_0})}{(c^*{}_0\varepsilon + c^*{}_0 - \frac{aN'}{y_0}) - a^*{}_0 \frac{x_0}{y_0}} , \tag{9}
$$

which is of the form

$$\frac{x}{y} = \frac{A\dfrac{x_0}{y_0} - B}{C - a^*_0\dfrac{x_0}{y_0}}. \tag{10}$$

Here $A > 0$ and $B > 0$. Since, with the above-mentioned assumption, x and y are both non-negative, therefore $C > 0$. When $x_0/y_0 = C/a^*_0$ then $x/y = \infty$, in other words $x = N'$, $y = 0$ and all of the passive population exhibit behavior A with an average intensity w_A. If y_0 is fixed, the requirement $x = N'$ gives

$$x_0 = Cy_0/a^*_0. \tag{11}$$

But C_0 is a linear function of $\varepsilon = \alpha_1 w_A$. Hence, the stronger the average intensity of activity A, the greater for a given y_0, must be x_0 in order to impress that activity on the whole passive population. If for a given x_0, w_A is too great, then the denominator of (10) will be positive and x/y will be finite, hence $x < N'$. If all the coefficients in our equations were known, then from a given maximum intensity w_A of behavior A which still can be imposed on *all* the passive individuals we could calculate x_0 for a known y_0.

We shall denote by w_{Am} the maximum value of w_A which still can be impressed on the whole passive population. Correspondingly we shall put $\varepsilon_m = \alpha w_{Am}$. For a given x_0 and y_0, ε_m is determined as the root of equation (11).

Let us now consider a simplified case, in which $w_B = 0$ and therefore $\varepsilon' = 0$. This means that group A merely resists the behavior A but does not tend to impose any qualitatively different behavior B. In that case

$$\frac{x}{y} = \frac{a^*_0\dfrac{x_0}{y_0} - \left(c^*_0 + \dfrac{aN'}{y_0}\right)}{\left(c^*_0\varepsilon + c^*_0 - \dfrac{aN'}{y_0}\right) - a^*\dfrac{x_0}{y_0}}. \tag{12}$$

In order to have $x = N'$, or $x/y = \infty$, we must have

$$a^*_0\frac{x_0}{y_0} + \frac{aN'}{y_0} = c^*_0(1 + \varepsilon_m). \tag{13}$$

Suppose now that the same group of x_0 active individuals tries to impose another behavior of intensity w_A on the passive population. Let this time that behavior A_1 be opposed by a different group of $y_{01} = \alpha y_0$

other active individuals, the coefficients of influence remaining the same. Then, denoting by x_1 and y_1 the number of passive individuals that correspondingly exhibit and do not exhibit behavior A_1, we have an expression similar to (12), in which ε_1 is put instead of ε, and αy_0 instead of y_0. Introducing (13) into that expression, we find, after elementary calculations:

$$\frac{y_1}{x_1 + y_1} = \frac{\alpha c^*_0(1 + \varepsilon_1) - c^*_0(1 + \varepsilon_m)}{\alpha c^*_0(1 + \varepsilon_1) - (\alpha c^*_0 + \dfrac{2aN'}{y_0})}, \tag{14}$$

which is of the form

$$\frac{y_1}{x_1 + y_1} = \frac{A' - B'\varepsilon_m}{C'}. \tag{15}$$

Let us discuss a case considered in a previous paper (3), namely that of two active classes I and II, and a passive class III. Let again class I represent the "controlling" or "governing class" and let x_0 refer to it. Class II again is the one that organizes the production of goods. We have discussed in *loc. cit.* some cases of interaction of two such classes, when class I requires a certain amount of goods produced by classes II and III to be surrendered to it. Here we shall consider the situation from a somewhat different point of view. Class I may require that every individual of class II and III gives a certain fraction of everything he produces to class I. Class II will oppose it, the opposition being the stronger, the greater the fraction required. Thus that fraction may be used as a measure of w_A. Class I will impose as high a w_{Am} as can be impressed on the whole population, and that fraction will be the larger, the larger x_0. In practice we may take as an illustration for w_{Am} the ratio of the governmental tax receipts to the total national income.

If we consider the case of very small a, then equations (10), (13), and (14) become simplified, their coefficient not containing y_0. In particular, because of (1), equation (13) now becomes of the form

$$\frac{x_0}{y_0} = A'' + B''w_{Am}, \tag{16}$$

while (15) becomes

$$\frac{y_1}{x_1 + y_1} = \frac{A''' - B'''w_{Am}}{C'''}. \tag{17}$$

Ascribing to w_{Am} the meaning above, we may try to check equation (16), if we have some other means of determining x_0/y_0. The following gives a very rough possible estimate of that ratio. Most of the

active population of a country is concentrated in cities. The stronger the "governing" class I, the more centralized the government and the larger the relative size of the capital city. Denoting by N_c the population of the capital and by N_u the total urban population, we may consider roughly

$$\frac{x_0}{y_0} = \frac{N_c}{N_u - N_c}. \tag{18}$$

Assuming that for different countries, the coefficients A'' and B'' are the same, which is of course only an extremely rough approximation, we shall expect

$$\frac{N_c}{N_u - N_c} \propto A'' + B'' w_{Am}. \tag{19}$$

Data of w_{Am} are scarce and inaccurate (5). Using what is available, the result of comparison of equation (19) with observation is shown in Figure 1, with data valid about 1930.

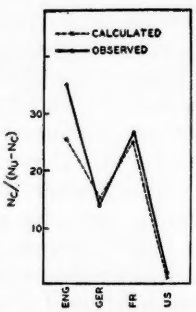

FIGURE 1

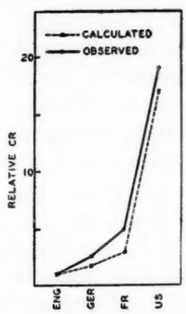

FIGURE 2

It must also be kept in mind that the whole discussion is based upon considerations of steady equilibrium states. A sudden increase of w_{Am} will not result in an immediate variation of $N_c/(N_u - N_c)$, according to equation (19). There will be considerable time lags governed by the general differential equation (4), in which w_A and hence ε are made explicit functions of time.

Equation (17) may be used for estimating the success of imposing other behaviors by x_0, in terms of w_{Am}. The governing class requires certain standards of behavior in different lines of life, the requirements being put into effect with different amounts of effort for different types of behavior. Deviations from the required behavior constitute crimes of various degrees. Thus equation (17) may be used

to study comparative criminality in different populations, for

$$y_1/(x_1 + y_1)$$

denotes the ratio of individuals that *do not* obey the dictates of class I. y_1 here denotes the total number of criminals, while y_{01} would denote the active ones. It will be agreed that a large number of crimes are committed by passive individuals, as a result of imitation, etc.

Criminal statistics are in themselves not very accurate and are hardly comparable for different countries due to different legal standards (5). The most comparable would be perhaps the incidence of such obviously criminal acts as murders. An important factor, however, has to be added in that case to equation (17). Other conditions, including x_0/y_{01}, being the same, there will be a difference in the incidence of crime depending on the ease with which a crime is or can be hidden. The latter depends on the density of population d. This function $f(d)$ of d is a rather complicated one. Obviously $f(d)$ must be zero for $d = 0$, since no crimes are committed in an unpopulated country. Yet beginning with rather small values of d, $f(d)$ must within a certain range of d's decrease with d, approximately as $1/d$. For the greater the density of population, the larger the number of individuals a crime affects, and the sooner it becomes known. A murder of an individual living alone in a secluded spot may remain undiscovered for weeks. As d increases further, however, this also increases the ease with which the criminal can hide himself after the deed. Thus $f(d)$ must start at zero, reach a maximum, decrease approximately as $1/d$, and then again increase. This latter increase is likely to be a factor of increased crime incidence in large cities, although a concentration of the active elements in cities, mentioned above, undoubtedly plays an important part, too. The density of population in large cities is of the order of 10^4 individuals per square kilometer or higher. The highest population densities in countries as a whole are about 300-400 individuals per square kilometer. Thus it is plausible to assume that the increase of $f(d)$ begins only over values of d of the order of 10^3 individuals per square kilometer. In identifying $y_1/(x_1 + y_1)$ with the incidence of murders CR, we will multiply the right side of (17) by $1/d$. The results given by the simple expression

$$CR = \frac{D - w_{Am}}{d}, \tag{20}$$

where D is a constant, are shown in comparison with observation in Figure 2.

We may consider a still different type of behavior, which is not forbidden by class I, but is not too much encouraged. As an example we may cite divorce. If $y_1/(x_1 + y_1)$, which we now identify with the divorce rate, *DR*, is very small, in other words $y_1 << x_1$, we may substitue for $y_1/(x_1 + y_1)$ simply y_1/x_1. For that ratio we have an expression of the same form as (10), namely,

$$\frac{y_1}{x_1} = \frac{\bar{A}\dfrac{y_{01}}{x_0} - \bar{B}}{\bar{C} - \dfrac{y_{01}}{x_0}}, \tag{21}$$

where again y_{01} is the number of active individuals advocating divorces. As before, we have $y_{01} = \alpha y_0$. When y_1/x_1 is small, then we may expand the right side of (21) and stop at the linear term. Thus we obtain

$$DR = \frac{y_1}{x_1} = \bar{A}\frac{y_0}{x_0} - \bar{B}'. \tag{22}$$

Combining (22) with (18) we find a relation between y_1/x_1 and $(N_u - N_c)/N_c$, which is illustrated in Figure 3.*

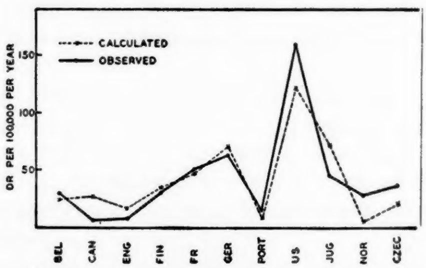

FIGURE 3

It must be emphasized that the foregoing illustrations do not mean any "confirmation" of a particular theory. They merely serve to illustrate how, starting from purely theoretical abstract concepts,

* In the illustration in Figure 1 of equation (19), W_{Am} for the United States was taken as the ratio of the total federal tax receipts to the national income. Accordingly, the population of Washington D. C. was taken for N_c. Considering that, due to the decentralized system of the United States Government, such regulations as concern divorces are more of a state nature, in the computations for Figure 3 by means of equations (22) and (18), N_c is taken as representing the sum of populations of all state capitals.

we may gradually arrive at relations that can be tested by observation. To speak of actual verifications of any such theory would require much more elaboration of the theory, which must take into account many more complex factors. The illustrations show, however, how certain relations may be suggested even by an inadequate theory, which thus helps us to notice such relations. For we usually notice only what we look for, and we look for things which we expect.

REFERENCES

1. Rashevsky, N. Studies in mathematical theory of human relations. *Psychometrika*, 1939, 4, 221-239.
2. Rashevsky, N. and Alston S. Householder. On the mutual influence of individuals in a social group. *Psychometrika*, 1941, 6, 317-321.
3. Rashevsky, N. Contributions to the mathematical theory of human relations. V. *Psychometrika*, 1942, 7, 117-134.
4. Rashevsky, N. Contributions to the mathematical theory of human relations: VI. Periodic fluctuations of the behavior of social groups. *Psychometrika*, in press.
5. Statistical Abstracts of the United States, 1939. Statistisches Jahrbuch für das Deutsche Reich, 1937, part II, internationale Ubersicht. Woytinsky, W., Die Welt in Zahlen; Mosse-Berlin. Snyder, Carl, Capitalism the creator; McMillan: New York, 1940.

PSYCHOMETRIKA—VOL. 7, NO. 4
DECEMBER, 1942

USE OF THE TEST SCORING MACHINE AND THE GRAPHIC ITEM COUNTER FOR STATISTICAL WORK

BENJAMIN S. BLOOM AND ARDIE LUBIN
BOARD OF EXAMINATIONS, THE UNIVERSITY OF CHICAGO

The graphic item-counter is described and its use as a statistical device is explained. Procedures are presented for obtaining Pearson product-moment correlations by means of the graphic item-counter.

The Test Scoring Machine developed by the International Business Machines Corporation is designed to secure by electrical means the score made by an individual who has marked an answer sheet according to the directions given in a test. A recently developed special device for this machine is the Graphic Item Counter. This attachment prints a graphic record of the pencil marks appearing in predetermined positions on a group of answer sheets. These graphs furnish the data necessary for item analysis, questionnaire analysis, and many requirements of response counting where original records may take the form of marks in particular positions on a machine-scored answer sheet.

The Graphic Item Counter has 90 counting positions. It is equipped with a plugboard which has one plugging position for each of the 750 response positions on the standard answer sheet or record form. It also has one plugging position for each of the 90 counting positions. Any response position may be connected to any counter by means of a plugwire.

To make an analysis of the marks on an answer sheet, the desired response positions are wired to the desired counters by means of the plugboard arrangement. The answer sheets are then passed through the machine. As each answer sheet passes through the machine, the marks are scanned and where a mark occurs in the proper position, the appropriate counter will register one. When the last answer sheet of the group has been passed through the machine, a graphic item count record sheet is inserted into the machine with a carbon paper over it. A starting lever is turned and the carriage automatically runs the sheet through the machine and prints on it a bar graph of the item count. The bars for each of the items project vertically so that the top mark in each column represents, by its position, the number

counted for that item. There are also two total counters which count the number of sheets run through the machine.

It is the purpose of this paper to describe a few of the possible uses of this machine for statistical work, with special reference to a method whereby the machine can be used for the computation of tables of intercorrelations. This machine can count as many as 90 marks at a time on a single answer sheet. Since about 500 papers can be passed through the machine in a single hour, it is possible to secure about 45,000 counts per hour. Such rapid counting should prove of great value in facilitating the computation of many of the statistical formulas common in education and psychology. This machine can be used for much of the statistical work which can be performed by the use of Hollerith or punch-card equipment.

Coding

An answer sheet may be used to record an individual's answers to a test or questionnaire. But, an answer sheet may also be used to record other types of information about a person or group. Since the answer sheet has 750 positions and each position may be used in the recording of information, a vast amount of information about a single individual or group may be placed on an answer sheet.

The following illustration may make clear some of the possibilities. If it is desired to indicate that John Doe is male, white, 14 years of age, and in the 8th year of school, and that he has a score of 30 on Arithmetic achievement, a score of 13 on Reading Aptitude, and an average grade of B, this may be recorded as shown in the illustration.

John Doe is a male, and since in our coding arrangement the male occupies the first response position on the answer sheet, that position has been blackened. Since John Doe is white, the first response position in the second row has been blackened. The geometric code is explained in the note. John's average grade is a B, and since B occupies the second position in the coding arrangement, the answer sheet has been blackened in the row corresponding to average grade as indicated above. Information about other individuals like John Doe might be coded in a similar fashion on other answer sheets.

After the information has been placed on the answer sheet, it is possible to pass the answer sheets through the machine and to determine the frequency with which each characteristic is present for an entire group of individuals. When a geometric or other code is used, it is possible to determine the frequency of each coded mark on the answer sheet and then to translate the coded marks back to the original characteristics.

Characteristic	Coding Arrangement	Answer Sheet Form
1. Sex	1. Male 2. Female	1
2. Color	2. White 2. Negro 3. Other	2
3. Age	Geometric code*	3
4. Arithmetic Achievement	" "	4
5. Reading Aptitude	" "	5
6. School Year	" "	6
7. Average grade	1. A 2. B 3. C 4. D 5. F	7

* Geometric code: Such a code utilizes a series such as 1, 2, 4, 8, 16, 32, etc. to represent a numerical value. Any value up to 31 may be represented by one number or by a combination of the numbers in the series 1,2,4,8,16. For example, the age 14 in the illustration above may be represented by the 2,4, and 8. We have used the five response positions of the answer sheet to indicate this series, so the age 14 was coded by blackening the 2nd, 3rd, and 4th positions which represent the 2nd, 3rd, and 4th numbers in the series. The value 30, which is the Arithmetic Achievement score, would be represented by the combination of 2,4,8, and 16, or the 2nd, 3rd, 4th, and 5th positions of the answer sheet. This geometric code and its uses have been treated in: Royer, E. B., and Toops, H. A., Statistics of geometrically coded scores, *J. Amer. Stat. Assoc.*, 1933, 28, 192-198.

If the individual answer sheets are sorted into groups on the basis of any desired characteristic such as sex, age, etc., a count and comparison can be made between the various groups for any data which have been coded onto the answer sheets.

Adding

Since any counting process is really an adding operation, this machine can be used for many tasks which involve addition. Adding may be done as follows:

1. If it is desired to add one column of marks to another, the length of each can be measured with a scale and the total length may be read in terms of the total number of marks or total frequency.

   Length of column_1 + length of column_2 = length of columns 1 + 2 = total number of marks in columns 1 and 2.

2. If the marks represent coded scores, the length of each column of marks on the graphic item count may be interpreted in terms of the appropriate unit or scale, and the summing of the desired columns will yield the total sum.

   Length of $\text{column}_1 \times$ appropriate unit + length of $\text{column}_2 \times$ appropriate unit = total sum.

Subtraction

The difference between the heights of two columns when read on a scale in terms of the appropriate unit will yield the result of the subtraction of one from the other.

Length of column_1 − length of column_2 = difference in number of marks in columns 1 and 2.

Multiplication

1. If it is desired to multiply one column by a certain number, a scale in terms of that unit may be placed beside the column and the height of the column when read on the scale will yield the desired sum.

   Length of $\text{column}_1 \times$ appropriate unit = product.

2. It is also possible by the proper wiring of the plugboard and the use of the Multiple Response Unit to determine the frequency with which one mark on the answer sheet occurs concurrently with another mark. The height of the appropriate column may be read on a scale marked in terms of $\text{Unit}_1 \times \text{Unit}_2$.

Tetrachoric correlations

The basic frequencies for a series of tetrachoric correlations may be obtained in the following ways:

1. The frequency with which A_1 and A_2 occur simultaneously may be obtained for as many as 90 interrelationships by wiring the plugboard so that A_1 and A_2 are wired to one counter, A_3 and A_4 are wired to another counter, etc., and the Multiple Response switch is turned on. A count will be made when each pair of marks occurs. When all the papers are passed through the machine, a graphic record will be produced, where each column represents the frequency with which each combination has occurred. Each column will represent the frequency in one cell of a four-fold distribution. The other cells can be determined by a knowledge of the totals for each row and column of the distribution. For example:

		2		
		A	B	
1	A	25		70
	B			30
		40	60	100

		4		
		A	B	
3	A	30		50
	B			50
		45	55	100

2. The answer sheets may be sorted into the A and B positions or piles for variable 1, and the frequency with which A_1 occurs with $A_2, A_3, A_4, \cdots, A_{90}$ may be determined by appropriate wiring and passing the answer sheets through the machine.

		2		
		A	B	
1	A	25		70
	B			30
		40	60	100

		3		
		A	B	
1	A	15		70
	B			30
		50	50	100

By the methods above, tetrachoric correlations may be computed between items, or between coded characteristics already placed on the answer sheet.

Product-Moment Correlations

Steps in Coding Scores

Product-moment correlations may be computed from data supplied by the Graphic Item Counter, when it is desired to secure all the intercorrelations for a large number of variables. The following steps may be followed:

1. Translate raw scores into step intervals. Sixteen or fewer step intervals represent a convenient division.
2. Convert step intervals to a geometric code using the numbers 1-2-4-8, e.g.:

   A score of 13 would be represented by the numbers 1-4-8;
   A score of 7 would be represented by the numbers 1-2-4;
   A score of 4 would be represented by the number 4.
3. For each individual use a 5-response answer sheet,

   letting response no. 1 = 1 of the code,
   letting response no. 2 = 2 of the code,
   letting response no. 3 = 4 of the code,
   letting response no. 4 = 8 of the code;

and allowing each question number to represent a variable, thus:

question no. 1 = variable 1,
question no. 2 = variable 2,
question no. 3 = variable 3.

Thus the scores of John Smith on a number of tests might be coded onto the answer sheet by the following steps:

Variable	Variable Number	Score	Step Interval	Code	Answer Sheet
Intelligence	1	86	15	1–2–4–8	
Arithmetic	2	29	7	1–2–4	
Spelling	3	54	11	1–2–8	
English	4	36	8	8	

Steps in Operation of Graphic Item Counter

4. Plug the Graphic Item Counter plugboard so that:

   Variable I—Response 1 is wired to counter 1,
   Response 2 is wired to counter 2,
   Response 3 is wired to counter 3,
   Response 4 is wired to counter 4.

   Variable II—Response 1 is wired to counter 5,
   Response 2 is wired to counter 6,
   Response 3 is wired to counter 7,
   Response 4 is wired to counter 8.

   Continue this until

   Variable XXII—Response 4 is wired to counter 88.

5. Divide the papers for the entire group of individuals on the basis of response positions blackened for Variable I. Select all answer sheets which have response No. 1 blackened for Variable I. This is done regardless of any other marks on this variable.

6. Run these papers through the graphic item counter and print the record which results from this process. Label this graph, Variable I, Code 1.

7. Place papers used in step 6 with other papers, and then select all answer sheets which have response No. 2 blackened for Variable I. This is done regardless of any other marks on this variable.

8. Repeat Step 6 for these papers and label the graph, Variable I, Code 2.
9. Repeat steps 7 and 8 for each response position on Variable I. Label results from response position 3, Variable I, Code 4. Label results from response position 4, Variable I, Code 8.
10. Repeat steps 5 through 9 for each variable, thus obtaining four graphs for each variable, one for each geometric code number used.

Steps in Securing Sums and Sums of Cross-Products From Graphic Item Counter Record

11. In a table similar to Table 1, copy the results from each graphic item counter record. Thus Variable I, Code 1 would be entered in row no. 1; Variable I, Code 2 would be entered in row no. 2. Continue this process until all rows have been filled. Thus each graph record corresponds to a row in this table.
12. Each entry in the cells should be multiplied by the row and column code numbers for which it forms the intersection. For example, this may be performed for Variable I by multiplying appropriate entries in the first row by the column code numbers 1, 2, 4, and 8. Enter the result in the appropriate cell in the column headed by a $\sum$ sign for Variable I. Repeat this operation for rows 2, 3, and 4. Then multiply the sum for row 1, Variable I by 1; row 2, Variable I by 2; row 3, Variable I by 4; and row 4, Variable I by 8. The resulting sum $= \sum X_1 X_1$. When these steps are repeated for Variables II and III, the results will equal $\Sigma X_1 X_2$ and $\Sigma X_1 X_3$. In this table the sums of the cross products have been indicated by encircling.*
13. To obtain ΣX for any variable, multiply the diagonal cell entries (the figures in parentheses) for each diagonal block by the appropriate column code numbers. The sum of the products for each diagonal block will be the ΣX for that variable. Thus in order to secure ΣX_1 in this example: $(1 \times 82) + (2 \times 105) + (4 \times 98) + (8 \times 91) = \Sigma X_1 = 1412$. These ΣX's have been placed at the top of the table.

Note: The correlations may be computed using any of the standard formulas for Pearson Product-Moment correlations, for example:

$$r = \frac{N\Sigma XY - \Sigma X \, \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \, \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}} \, .$$

TABLE 1
Crossproduct Matrix Utilizing a Geometric Code

Graphic Item Counter Numbers			$\Sigma X_1 = 1412$					$\Sigma X_2 = 1282$					$\Sigma X_3 = 1921$				
			1	2	3	4		5	6	7	8		9	10	11	12	
			Variable I					Variable II					Variable III				
Row No.	Variable Number	Code No.	1	2	4	8	Σ	1	2	4	8	Σ	1	2	4	8	Σ
1		1	(82)	42	37	41	642	34	44	41	28	510	6	45	47	58	748
2		2	42	(105)	50	56	900	51	53	59	38	697	11	59	65	80	1029
3	Variable I	4	37	50	(98)	66	1057	43	48	54	38	659	11	54	53	68	875
4		8	41	56	66	(91)	1145	44	44	45	53	736	2	45	46	56	724
							$\Sigma X_1X_1 = 15{,}830$					$\Sigma X_1X_2 = 10{,}428$					$\Sigma X_1X_3 = 12{,}098$
5		1	34	51	43	44	660	(92)	41	49	40	690	7	45	49	70	853
6		2	44	53	48	44	694	41	(101)	53	41	783	10	62	61	78	1002
7	Variable II	4	41	59	54	45	735	49	53	(109)	43	935	17	63	71	84	1099
8		8	28	38	38	53	680	40	41	43	(69)	846	1	33	32	40	515
							$\Sigma X_2X_1 = 10{,}428$					$\Sigma X_2X_2 = 12{,}764$					$\Sigma X_2X_3 = 11{,}373$
9		1	6	11	11	2	88	7	10	17	1	103	(19)	19	19	19	285
10		2	45	59	54	45	739	45	62	63	33	685	19	(115)	68	86	1209
11	Variable III	4	47	65	53	46	757	49	61	71	32	711	19	68	(122)	90	1363
12		8	58	80	68	56	938	70	78	84	40	882	19	86	90	(148)	1735
							$\Sigma X_3X_1 = 12{,}098$					$\Sigma X_3X_2 = 11{,}373$					$\Sigma X_3X_3 = 22{,}035$

The column numbers correspond to the counter-numbers which were plugged to the appropriate variable and response number. Except for the diagonals, the cell entries in each column will be found to have duplicates in the corresponding row positions. Thus the table will be symmetrical about the diagonal running from upper left to lower right.

By this method it is possible for two clerks to compute 231 intercorrelations using 100 cases in about four working days, with every correlation computed twice using different basic figures. Accurate results will be obtained if the original answer sheets are marked very carefully with heavy marks and all stray pencil marks are removed.

* An alternative method for transforming geometric coded values to sums of cross products is given by Kuder, G. F. Use of the International Scoring Machine for the rapid computation of tables of intercorrelations, *J. appl. Psychol*, 1938, 587-596.

PSYCHOMETRIKA—VOL. 7, NO. 4
DECEMBER, 1942

ON DETERMINING THE RELIABILITY AND SIGNIFICANCE OF A TETRACHORIC COEFFICIENT OF CORRELATION*

J. P. GUILFORD AND THOBURN C. LYONS
UNIVERSITY OF SOUTHERN CALIFORNIA

In this note are presented facilitating tables for the estimation of the standard error of a tetrachoric r and also tables providing significant and very significant tetrachoric coefficients for various sizes of samples and various combinations of proportions in the dichotomized distributions.

The tetrachoric coefficient of correlation has been coming more and more into use in recent years, particularly since the Thurstone computing diagrams (1) are generally available. There is reason to believe that the popularity of this statistic will continue. It is important, therefore, that consideration be given to the question of the reliability and the statistical significance of the tetrachoric r.

The complete formula for estimating the standard error of this kind of coefficient is so forbidding in terms of labor that rarely does a textbook on statistics present it. And yet, because the standard error is so much larger than that for an ordinary Pearson r under similar circumstances, it is important that the research worker be aware of its magnitude when he computes a tetrachoric r.

The present trend in sampling theory is to use the standard error rather than the probable error of an estimated parameter, so that practice will be observed here. According to Kelley (2), the standard error of a tetrachoric r is given by the formula

$$\sigma_r = \frac{\sqrt{pp'qq'}}{yy'\sqrt{N}} \sqrt{\left[1 - \left(\frac{\sin^{-1} r}{90^\circ}\right)^2\right](1 - r^2)}\,, \tag{1}$$

in which p is the proportion of the cases in one of the two main categories for one of the correlated variables,

p' is the similar proportion for the other variable,

$q = 1 - p$, and $q' = 1 - p'$,

y and y' are ordinates in the normal distributions of unit

* The task of computing the values in the accompanying tables should be credited to Mr. Lyons.

area at the deviates which correspond to p and p' respectively,

r is the tetrachoric coefficient of correlation,

and $\sin^{-1}r$ is the angle whose sin is equal to r .*

For convenience in what follows we have envisaged this formula as being factored as follows:

$$\sigma_r = \frac{1}{\sqrt{N}} \cdot \frac{\sqrt{pq}}{y} \cdot \frac{\sqrt{p'q'}}{y'} \cdot \sqrt{\left[1 - \left(\frac{\sin^{-1}r}{90^\circ}\right)^2\right] \cdot (1 - r^2)} \,. \tag{2}$$

Let us call the five factors A , B , C , D , and E , respectively. The equation then reads

$$\sigma_r = A \cdot B \cdot C \cdot \sqrt{D \cdot E} \,. \tag{3}$$

It is our purpose to present, first, tabled values for factors B , C , and $\sqrt{DE}$ which will facilitate decidedly the computation of σ_r . Similar tables for this purpose have appeared before, but not collectively or in complete form, and only then for the purpose of estimating PE_r rather than σ_r (2,3). Table 1 gives values for both B and C . Entering this Table with either p or q (whichever is .50 or larger) and then with p' or q', we can read values of B and C . Table 2 provides values for the factor $\sqrt{DE}$ for values of r ranging from 0.00 to 0.99 in steps of 0.01. Factor A is readily determined from general tables of square roots and reciprocals or by computation. The product of the four factors yields σ_r . The use and interpretation of this estimated parameter is of course subject to the same restrictions and qualifications as in the case of any σ_r .

Probably of greater utility in interpreting a coefficient of correlation is the practice of determining whether an r is far enough removed from zero to be indicative of a genuine correlation in the population from which the sample was drawn. The practice of computing σ_r , which was discussed above, is based upon the assumption that the true r , or population r , is identical with the sample r . Furthermore, as Fisher has often pointed out, the distribution of the sample r's is skewed when r is large so that the usual interpretations of the fluctuations of sample r's are sometimes most unsatisfactory. To the knowledge of the writers, there is no provision for translating a σ of a tetrachoric r into terms of a z parameter which is symmetrically distributed, as is true for the ordinary Pearson r . The best solution, therefore, seems to be the assumption of a null hypothesis. This means to

* A misprint appearing in Kelley's presentation of the formula has been corrected here.

suppose that the population correlation is actually zero and to compute σ_r to fit this assumption. A distribution of such sample r's would be symmetrical, and from the size of this σ_r and of the obtained coefficient we can infer the probability that the null hypothesis is tenable or untenable.

In line with this discussion, we have adopted Fisher's fiducial limits of 5 per cent and 1 per cent and Student's distribution as bases of deciding whether a certain tetrachoric r is significant or very significant. An r is regarded as significant if for a sample of size N there is only 1 chance in 20 of obtaining an r as large or larger in random sampling from the same population. An r is regarded as very significant if there is only 1 chance in 100 of obtaining similarly an r that deviates that much or more from zero. We present in Table 3 the significant tetrachoric r's for various combinations of N and of p and p'. We present in Table 4, similarly, the very significant tetrachoric r's. To be specific, when N is 100, when p (or p') is .6, and p' (or p) is .5, it takes an r of at least 0.315 to be regarded as significant (see Table 3). An r as large as 0.315 or larger, either positive or negative, could occur simply by random sampling in an uncorrelated population 5 times in 100, when the size of sample is 100. For the same population and size of sample, Table 4 tells us that it would take an r of 0.417 to be regarded as very significant. Once in a hundred times a tetrachoric r as large as this or larger could occur when the true correlation is zero.

In using Tables 3 and 4, there is a general rule that p and p' are interchangeable. To take an example, from the column headed $p = .9$ and $p' = .6$, one can also find the significant (or very significant) r for the case in which $p = .6$ and $p' = .9$. Assume that $N = 250$, $p = .9$ and $p' = .6$, and a significant r is 0.270 and a very significant one 0.356. The same values of r would apply when $p = .6$ and $p' = .9$.

Another rule is that p and q are interchangeable. When p is less than .5, one must enter the table with q, which equals $1 - p$. For example, if the obtained p is .2 and p' is .4, in the table we look for $p = .8$ and $p' = .6$. To take another example, if p and p' both equal .1, we look in the column headed $p = .9$ and $p' = .9$. This type of replacement also holds when only one p is less than .5. For the combination $p = .3$ and $p' = .8$, we would look for the heading $p = .7$ and $p' = .8$. But since p is always greater than or equal to p' in these particular tables, we look for $p = .8$ and $p' = .7$.

If the obtained proportions and values of N do not coincide with those offered in the tables, one may perform the necessary interpolations. It is doubtful whether the labor of interpolating is worth while

except when the obtained r is quite near the boundary line of significance, however. In other instances, one might be conservative by taking the next smaller N than his sample contained, and by choosing p values nearer to 1.00 than the obtained ones.

Inspection of Tables 3 and 4 shows that within their limits the significant r's range from 0.097 to 0.580, and very significant r's range from 0.128 to 0.767. These facts should impress one with the importance of working only with very large samples when a tetrachoric r is to be the index of correlation. Only then will σ_r be reasonably small and will one be justified in rejecting the null hypothesis when r turns out to be small or even moderate in size.

REFERENCES

1. Chesire, L., Saffir, M., and Thurstone, L. L. *Computing Diagrams for the Tetrachoric Correlation Coefficient.* Chicago: Univ. Chicago Press, 1933.
2. Kelley, T. L. *Statistical Method.* New York: Macmillan Company, 1924.
3. Davenport, C. B. and Ekas, M. P. *Statistical Methods in Biology, Medicine, and Psychology.* New York: John Wiley & Sons, 1936.

TABLE 1
Providing the Values for Factors B and C in Formula (2) Corresponding to Various Values of p or q

p or q	$\frac{\sqrt{pq}}{y}$	p or q	$\frac{\sqrt{pq}}{y}$	p or q	$\frac{\sqrt{pq}}{y}$	p or q	$\frac{\sqrt{pq}}{y}$	p or q	$\frac{\sqrt{pq}}{y}$
.50	1.2533	.60	1.2680	.70	1.3180	.80	1.4287	.90	1.7094
.51	1.2535	.61	1.2712	.71	1.3256	.81	1.4457	.91	1.7623
.52	1.2539	.62	1.2748	.72	1.3338	.82	1.4641	.92	1.8248
.53	1.2546	.63	1.2787	.73	1.3427	.83	1.4844	.93	1.9003
.54	1.2556	.64	1.2830	.74	1.3523	.84	1.5067	.94	1.9936
.55	1.2569	.65	1.2877	.75	1.3626	.85	1.5315	.95	2.1131
.56	1.2585	.66	1.2928	.76	1.3738	.86	1.5590		
.57	1.2604	.67	1.2984	.77	1.3859	.87	1.5897		
.58	1.2626	.68	1.3044	.78	1.3990	.88	1.6245		
.59	1.2652	.69	1.3109	.79	1.4133	.89	1.6640		

TABLE 2
Providing the Values for the Factor $\sqrt{DE}$ in Equation (2) Corresponding to Different Values of the Tetrachoric r

r	$\sqrt{DE}$	r	$\sqrt{DE}$	r	$\sqrt{DE}$	r	$\sqrt{DE}$	r	$\sqrt{DE}$
.00	1.0000	.20	.9717	.40	.8845	.60	.7297	.80	.4844
.01	.9998	.21	.9687	.41	.8784	.61	.7199	.81	.4686
.02	.9997	.22	.9657	.42	.8723	.62	.7099	.82	.4526
.03	.9994	.23	.9625	.43	.8658	.63	.6996	.83	.4362
.04	.9988	.24	.9591	.44	.8594	.64	.6892	.84	.4191
.05	.9982	.25	.9555	.45	.8526	.65	.6784	.85	.4018
.06	.9975	.26	.9520	.46	.8458	.66	.6675	.86	.3838
.07	.9966	.27	.9483	.47	.8388	.67	.6563	.87	.3652
.08	.9955	.28	.9442	.48	.8314	.68	.6448	.88	.3460
.09	.9942	.29	.9401	.49	.8240	.69	.6331	.89	.3262
.10	.9930	.30	.9358	.50	.8165	.70	.6210	.90	.3057
.11	.9915	.31	.9314	.51	.8087	.71	.6087	.91	.2844
.12	.9899	.32	.9268	.52	.8007	.72	.5961	.92	.2620
.13	.9881	.33	.9220	.53	.7926	.73	.5834	.93	.2387
.14	.9862	.34	.9171	.54	.7841	.74	.5702	.94	.2142
.15	.9841	.35	.9122	.55	.7755	.75	.5569	.95	.1881
.16	.9819	.36	.9070	.56	.7669	.76	.5429	.96	.1606
.17	.9795	.37	.9016	.57	.7579	.77	.5288	.97	.1305
.18	.9770	.38	.8961	.58	.7488	.78	.5145	.98	.0973
.19	.9745	.39	.8904	.59	.7394	.79	.4995	.99	.0586

TABLE 3

Tetrachoric Coefficients of Correlation, Significant at the .05 Level, for Various Sizes of Sample and Combinations of p and p'

N	$p=$ .9 $p'=$ .9	.9 .8	.9 .7	.9 .6	.9 .5	.8 .8	.8 .7
100	.580	.485	.447	.430	.425	.405	.374
150	.471	.394	.363	.350	.346	.329	.304
200	.407	.340	.314	.302	.299	.285	.263
250	.364	.304	.281	.270	.267	.254	.235
300	.332	.277	.256	.246	.243	.232	.214
350	.307	.257	.237	.228	.225	.215	.198
400	.287	.240	.222	.213	.211	.201	.185
500	.257	.215	.198	.191	.188	.179	.166
600	.234	.196	.181	.174	.172	.164	.151
800	.203	.170	.156	.151	.149	.142	.131
1000	.181	.151	.140	.134	.133	.127	.117
1500	.148	.124	.114	.110	.108	.103	.095
2000	.128	.107	.099	.095	.094	.089	.083
2500	.115	.096	.088	.085	.084	.080	.074
3000	.105	.087	.081	.078	.077	.073	.067
5000	.081	.068	.062	.060	.059	.057	.052
10000	.057	.048	.044	.043	.042	.040	.037

TABLE 3 (continued)

N	$p=$ .8 $p'=$ .6	.8 .5	.7 .7	.7 .6	.7 .5	.6 .6	.6 .5	.5 .5
100	.360	.355	.345	.332	.328	.319	.315	.312
150	.292	.289	.280	.270	.267	.259	.256	.253
200	.253	.250	.242	.233	.230	.224	.222	.219
250	.226	.223	.216	.208	.206	.200	.198	.196
300	.206	.203	.197	.190	.188	.183	.181	.179
350	.190	.188	.183	.176	.174	.169	.167	.165
400	.178	.176	.171	.164	.162	.158	.156	.154
500	.159	.157	.153	.147	.145	.141	.140	.138
600	.145	.144	.139	.134	.133	.129	.128	.126
800	.126	.124	.121	.116	.115	.112	.110	.109
1000	.112	.111	.108	.104	.102	.100	.099	.097
1500	.092	.091	.088	.085	.084	.081	.080	.080
2000	.079	.079	.076	.073	.072	.071	.070	.069
2500	.071	.070	.068	.066	.065	.063	.062	.062
3000	.065	.064	.062	.060	.059	.058	.057	.056
5000	.050	.050	.048	.046	.046	.045	.044	.044
10000	.036	.035	.034	.033	.032	.032	.031	.031

TABLE 4

Tetrachoric Coefficients of Correlation, Significant at the .01 Level, for Various Sizes of Sample and Combinations of p and p'

N	$p=$ .9 $p'=$ .9	.9 .8	.9 .7	.9 .6	.9 .5	.8 .8	.8 .7
100	.767	.641	.592	.569	.563	.536	.494
150	.622	.520	.480	.462	.456	.435	.401
200	.537	.449	.414	.399	.394	.375	.346
250	.480	.401	.370	.356	.352	.335	.309
300	.437	.365	.337	.324	.321	.305	.282
350	.404	.338	.312	.300	.297	.283	.261
400	.378	.316	.292	.281	.277	.264	.244
500	.338	.282	.260	.251	.248	.236	.218
600	.308	.258	.238	.229	.226	.215	.199
800	.267	.223	.206	.198	.195	.186	.172
1000	.238	.199	.184	.177	.175	.166	.153
1500	.194	.163	.150	.144	.143	.136	.125
2000	.168	.141	.130	.125	.123	.118	.109
2500	.151	.126	.116	.112	.110	.105	.097
3000	.137	.115	.106	.102	.101	.096	.089
5000	.106	.089	.082	.079	.078	0.74	.069
10000	.075	.063	.058	.056	.055	.053	.049

TABLE 4 (continued)

N	$p=$ .8 $p'=$ .6	.8 .5	.7 .7	.7 .6	.7 .5	.6 .6	.6 .5	.5 .5
100	.476	.470	.456	.439	.434	.422	.417	.413
150	.386	.381	.370	.356	.352	.343	.339	.335
200	.333	.329	.320	.307	.304	.296	.292	.289
250	.298	.294	.285	.275	.271	.264	.261	.258
300	.271	.268	.260	.250	.247	.241	.238	.235
350	.251	.248	.240	.231	.229	.223	.220	.217
400	.234	.232	.225	.216	.214	.208	.206	.203
500	.209	.207	.201	.193	.191	.186	.184	.182
600	.191	.189	.183	.176	.174	.170	.168	.166
800	.165	.163	.158	.152	.151	.147	.145	.143
1000	.148	.146	.142	.136	.135	.131	.130	.128
1500	.121	.119	.116	.111	.110	.107	.106	.105
2000	.104	.103	.100	.096	.095	.093	.091	.091
2500	.093	.092	.099	.086	.085	.083	.082	.081
3000	.085	.084	.082	.079	.078	.076	.075	.074
5000	.066	.065	.063	.061	.060	.059	.058	.057
10000	.047	.046	.045	.043	.043	.041	.041	.040

PSYCHOMETRIKA—VOL. 7, NO. 4
DECEMBER, 1942

A FACTORIAL STUDY OF AUDITORY FUNCTION

J. E. KARLIN
UNIVERSITY OF CHICAGO

Tests of auditory function in the fields of pitch, loudness, quality (timbre), and time, auditory analysis, synthesis, and memory, together with age, intelligence, and four tests of visual memory, were studied factorially. The subjects were 200 high-school students. The intercorrelations were factored to nine factors by a modification of the centroid technique and rotated to an oblique simple structure. No general auditory factor appeared. Instead there appeared group factors tentatively identified as pitch-quality discrimination, loudness discrimination, "auditory integral for perceptual mass," auditory resistance (synthesis and analysis), speed of closure, auditory span formation, memory span (auditory and visual), memory or incidental closure and an unidentifiable residual plane. The average intercorrelation among the primary vectors was low, only one intercorrelation being greater than .34. A number of queries are answered by the interpretation of the results.

I. Statement of the Problem

Preliminary factorial investigation of parts of the auditory field (6) has indicated that individual differences in auditory proficiency can reasonably be supposed to arise from a structured matrix of fundamental abilities. Although previous factorial studies have been exploratory in purpose and small in extent, a satisfying amount of agreement has been demonstrated among the different studies. It was decided, therefore, to investigate the auditory field in greater detail; the result is the present study. The line of attack was derived mainly from two sources, (a) leads from previous factorial work and (b) experimental and clinical evidence.

It has been established in previous analyses (6) that even when as few as six auditory tests were considered there was no general auditory factor. The low communalities of the tests had argued for a complex functional background. The evidence was thus all against the conventional clinical assumption that there was a strong general factor operative in various types of auditory situations; this factor was measured by an audiometer test. The clinical experience of the present writer had shown, furthermore, that except for cases of extreme deterioration of the auditory apparatus there frequently existed significant discrepancies between audiometric measurements

and success of prediction of hearing of the spoken voice and other complex sounds in everyday acoustic situations. Two patients might have about the same *auditory acuity* as indicated by the audiometer but differ appreciably in their *auditory ability* as shown by their performance in a more complex auditory environments. The initial facets of the problem become then:

1. Would a more extensive factorial investigation of auditory phenomena verify the failure to discover a general auditory factor in earlier exploratory minor studies?

2. Is there any broad group factor which might be deemed a fair approximation to the general factor?

3. To what extent does pitch and loudness discriminatory sensitivity predict response to social auditory stimuli?

Together with the general factor assumption in audition, there has been a large amount of experimental activity based on what might be called the group-factor assumption in audition. This work has been carried out by the physicist investigating the properties of sound, the psycho-physiologist interested in general audition, the psychologist examining the auditory background of musical phenomena, and the clinical otologist furthering research in diagnosis of auditory pathology. For some idea of the prevailing evidence and theories concerning these types of auditory function reference should be made to the more modern texts in the field **(1, 4, 7, 11)**.

All experimentalists agree in taking their start from the physical characteristics of the sound wave. The assumption is made that the physical characteristics of frequency, intensity, complexity, and duration have four functionally-distinct corresponding types of auditory function in sense of pitch, sense of loudness, sense of timbre or quality, and sense of time. This assumption does not appear to have been questioned either on theoretical or empirical grounds.

Each of these four functionally distinct qualitative types of auditory function has been investigated experimentally with some thoroughness. The manner of function of each of these variables has been described in some detail, both when the other three variables are held constant and when they are allowed to vary. In general, it is found that each function does not remain constant when the other three variables are influential. It should be noted that these four factors have not been *demonstrated* with any degree of rigor to be functionally distinct; they have been *assumed* distinct on the grounds that their physical counterparts are theoretically separable in the mathematical analysis of the sound wave. The psychology of hearing has in fact developed along rather unpsychological lines. Customarily, new

ideas in psychology spring from observation of behavior; only if known psychological techniques are inadequate for the furtherance of these ideas does it become necessary to seek methods of approach from other sciences. Here the process tended to be reversed. As a result there exists at the present time a substantial auditory literature on various so-called psychological entities with a foundation which is physical rather than psychological.

Notwithstanding the great weight of experimentation on these four postulated primary auditory factors, it becomes apparent upon investigation that this experimental work has itself been based upon an increasing number of further assumptions. Each factorial name has been applied to a further series of auditory phenomena on theoretical grounds. For instance, first the assumption is made that there is a pitch factor for pure tones; then the term pitch factor is applied to complex sounds, short sounds, vocal sounds, and so on. The same is true for sense of loudness. The relation between the various sub-functions of each major type of auditory function has not been demonstrated.

At this point it is possible to mention some further aspects of the general problem in addition to the three already mentioned:

4. Are there four distinct functional unities of the character of pitch, loudness, quality, and time?

5. Can it be demonstrated that the various sub-functions of the categories pitch, loudness, quality, and time are sufficiently saturated with the dominant trait of that category to warrant the same psychological description for their essential character?

The first step towards setting up the battery of auditory tests in accordance with the line of reasoning outlined above was to allow for the four domains in the auditory field demanded by tradition. In a factor analysis it is necessary to construct at least two tests in order to stabilize a factor in the common factor space. A number of tests were therefore constructed for each domain; such a procedure would appear to comply with the prerequisites for the investigation of aspects 4 and 5 above. These domains are discussed in turn.

Pitch Domain

In the pitch domain were tests of pitch discrimination for pure tones (Test 1), for complex sounds (Test 2), for short-impulse pure tones (Test 3), and for vocal sounds (Test 4). It has been assumed that these are all based upon the same fundamental pitch function, although the empirical evidence for this assumption has not been forthcoming. The pitch of complex sounds has been difficult to meas-

ure either physically or psychophysically. In the present study an assumption has been made which has been shown valid in other branches of test-theory where a test is measuring systematic variability within the range of ability of the subjects, namely: where two complex sounds do not differ in those characteristics which make it possible to assign them definite pitches, the judgments of a large enough random sample from the general population will be about equally divided in the comparison of the two pitches. Where the judgments of such a sample show a significant majority for a pitch difference, the judgment of the majority can effectually be used as the correct response. The scoring criterion for pitch discrimination for complex sounds was therefore taken as the significant judgment of the majority for each comparison item. At worst, if this scoring device is invalid for co-relating systematic variabilities, the test would not show up in the factorial domain and some other method would have to be devised to score such a test. To the extent that the test does correlate with other tests and does appear in the factorial framework, it must be conceded that this scoring device is valid.

In the case of vocal sound pitch discrimination the scoring criterion was based on the experimental and psychophysical findings. In the study of speech dynamics it has been established that two vocal sounds may have the same fundamental frequency and yet be judged to possess different pitches. The conclusion has been that such differences are attributable to the complexity of overtones of the sounds. It would seem therefore that this test is equally well to be considered a test of quality discrimination and should have some projection in the quality domain.

In the short-impulse pitch discrimination test it was considered of interest to discover what relationship there might exist between the duration threshold necessary for accurate pitch judgments and other forms of pitch judgments above the duration threshold. The corresponding problem was likewise investigated in the loudness domain.

Loudness Domain

This domain was determined by tests of loudness discrimination for pure tones (Test 5), for complex sounds (Test 6), for short-impulse pure tones (Test 7), and for the pitch-loudness function (Test 8). Again it has been assumed that these are various aspects of the same basic loudness function and again the empirical verification is lacking. The difficulties in scoring a test of loudness discrimination for complex sounds are similar to those for complex pitch, especially as the complex stimuli are not sustained. The judgments of the majority were again taken as indicating the correct responses. The

pitch-loudness function is a psycho-physiological phenomenon and has no known physical explanation. The function is plotted as the relation of intensity threshold to frequency of pure tone stimuli. It turns out to be a function such that the extreme frequencies, high and low, require greater physical intensity to become audible (3, 12). In the present test loudness judgments are required for comparison of two tones of different frequency but the same intensity. It becomes of interest to discover whether individual differences in this function are functionally correlated with loudness differences or with pitch differences. Such a finding would be of immediate anatomical interest.

Quality Domain

It has already been hypothesized that the test of vocal pitch discrimination (Test 4) will have a projection in this domain. Any timbre or quality factor might reasonably be supposed to acquire further stabilizing projections from a number of tests in which the stimuli are complex sounds. In particular the maximum projection would apparently come from the test for quality discrimination (Test 13).

Time Domain

The conventional distinction has been made between "filled time" and "unfilled time." There does not appear to be any evidence on the underlying relation of these two aspects of the time sense. If there is a distinct factor analogous to sense of time, the two time tests (Test 9 and 10) employed here should bring it to light.

The remainder of the problem emerges directly from the question: "To what extent may the results of laboratory tests of these four categories, employing relatively simple and meaningless stimuli, be deemed predictive of auditory behavior response in the more complex and meaningful social situations of spoken and musical sound?" As has been previously pointed out, this seemed to be the problem that clinicians had not as yet solved.

The definition of the content and boundaries of the social auditory situation was the first source of concern in attempting a solution of this aspect of the problem. It was necessary to review the auditory literature and to set up tests which could be considered a useful sample of the types of auditory function common to the conventional auditory social environment. The further auditory domains finally chosen with this condition in mind were: Rhythm, Auditory Analysis, Auditory Synthesis, and Auditory Memory.

Rhythm Domain

The rhythm factor is obviously important in a variety of auditory situations; specifically, mention might be made of the role

of rhythm in speech pathology and in the appreciation of musical progressions. An adequate picture of the past and present status of experimental work on rhythm may be obtained from a study of the publications of the University of Iowa group specializing in the psychology of music. The over-all view expressed there (8) is that the rhythmic sense "consists essentially in a tendency to group a succession of auditory stimuli according to the relevant dynamics of time and stress." Factorially, this might be expressed in the view that rhythm is a function of a time factor, a loudness factor, and specific factors. In the present case two tests of rhythm were used: motor rhythm (Test 11) and music rhythm (Test 12).

Domain of Auditory Analysis

By auditory analysis is meant the power of the auditory mechanism to receive composite stimulation and to break down this complex sound into its component parts. In everyday life it is obvious that the ear is never subjected to any single, detached type of auditory stimulation. At any given time the ear is being literally bombarded by a multifarious array of sounds; if all such stimuli could claim attention in proportion to their physical energy the auditory environment would be in a chaotic state. Orderliness and meaningfulness are imposed by the selective power of the auditory centers so as to magnify psychologically the stimulating force of certain stimuli and diminish that of others in a manner somewhat independent of their physical energies. Thus normally only certain selected stimuli receive attention in consciousness. This problem of selection of auditory stimuli which are continually being obscured by other stimuli is typically investigated under the name of *masking*. By all accounts the field of masked phenomena is exceedingly intricate and the laws operating therein are a long way from being understood with any degree of comprehensiveness. The degree of factorial complexity of this domain and its relation to other auditory domains is still very much a matter for future investigation.

It was hoped that this part of the auditory field would draw to it the projections of at least Tests 14-17. All these tests presented auditory tasks requiring the selection of given stimuli from complex auditory situations; the complexity would appear to be in large part a function of the different kinds of masking used. In Test 14 (Sound Breakdown) five simultaneous isolated voice-sounds were masking each other; in Test 15 (Pure Tone Masking) one pure tone was masking a second pure tone; in Test 16 (Sensory Masking) the spoken voice was being masked by a continuous buzzing noise; in Test 17 (Intellective Masking) one intermittent voice was being masked by a

second continuous voice. One essential difference between Tests 16 and 17 would appear to be that in Test 16 the distracting sound is operative on the sensory level and has no meaning on higher levels; in Test 17 the distracting voice not only obscures the stimulus value of the primary voice but also competes for attention on more meaningful levels, especially since the content of the distracting speech is much more interesting than the content of the primary speech.

Domain of Auditory Synthesis

Distortion of the meaningfulness of sound in various contexts can occur also by processes other than those of masking. The distortion may arise from defects in the articulation of the vocal sound itself even when the environmental conditions are otherwise favorable for auditory perception. The function of the ear permitting reception and comprehension of auditory stimuli in spite of articulatory defects is implied in the term auditory synthesis. This may be defined as the ability to resist and compensate for sound temporally distorted. This type of distortion may take several forms:

1. Alteration of the habitual rate of presentation of vocal symbols tends to nullify meaning. If vocal sounds follow one another in too rapid sequence, the auditory mechanism may be stimulated on a point for point basis but the message carried by the auditory tracts may remain uninterpretable at the center. Test 19 (Rapid Spelling) was constructed to illustrate this function. The letters of common words are spelled out rapidly, much in the manner adopted by two adults wishing to converse on topics not intended for the enlightenment of the younger family members present. In such cases much difficulty is often experienced by the other adult in understanding the word spelled out.

2. Intelligible speech requires a certain minimum standard of conventional modes of articulation. Undue changes in the pitch, loudness, duration, inflection, and similar factors exert inhibitory influences against word-perception. Test 20 (Singing) and Test 21 (Haphazard Speech) appear to feature such factors. It is a familiar observation that the words of a song are much more difficult to understand than the words of spoken speech. The artistic demands of the song-form require a dynamic system of the foregoing factors in accordance with the nuances of the music and pay little attention to the intrinsic meaningfulness of the words themselves. The aesthetic qualities of the musical sounds are, furthermore, usually the more interesting to the listener and serve to raise the limen for verbal intelligibility. In Haphazard Speech the words of simple sentences are spoken with unconventional changes of the sort just mentioned. The

general effect is that of an extremely nervous public speaker who is in addition the unfortunate possessor of uncontrollable vocal cords.

3. The rhythmic arrangement of words in meaningful phrases is a *sine qua non* of intelligible speech. Each idea or phrase automatically forms an auditory gestalt; destruction of such gestalts within the sentence rends the sentence void of utility. In Test 23 (Illogically Grouping) the phrase-gestalten were purposefully altered so as to necessitate the reconstructive power of the auditory process before words achieved their customary form.

Domain of Auditory Memory

It seems fairly clear that much of the meaning derived from auditory stimuli is normally possible only because of the ability of the organism to retain in memory a succession of such stimuli for suitable interpretation. This ability is usually termed auditory memory. The amount of experimental literature reported on this topic is large but the quantity of evidence on the correlation of individual differences in this connection is much smaller. The validity of even the small amount of the latter type of evidence is questionable. However, it was found in the preliminary studies which led the way to the present investigation (6) that memory tended to play a rather surprisingly important part factorially in auditory functions.

Four tests of auditory memory were specifically designed to augment the auditory memory common factor variance and to provide a framework for possible memory factors. These tests were: Memory for Female Voices (Test 24), Memory for Male Voices (Test 25), Tonal Memory (Test 26), and Memory for Emphasis (Test 27).

The relation between auditory memory span and visual memory span is still controversial in spite of a great deal of evidence on this point. This lack of agreement might be due in part to the different contexts in which the span tests appeared. The factorial approach of a study such as the present one would at least indicate something of the nature of the relationship in so far as this relationship was based upon the factors sampled in the study. The more complete nature of the relationship would depend upon successive variations of the factorial context. In the present study, auditory memory span (Test 23) and visual memory span (Test 28) were used as a beginning towards the solution of this problem.

At this point the formulation and discussion of the auditory problem is really complete for present purposes. It was decided, however, to provide opportunity for evidence on two further subsidiary aspects of the same problem, namely, reading and speech disabilities, and visual memory functions. In some reading and speech disability cases

it is possible to demonstrate gross impairment of hearing; in other cases hearing for all frequencies appears normal but the sounds of different vowels and consonants cannot be discriminated. It was thought that this latter type of difficulty might be due to a defect of some more complex auditory function. Test 18 (Vowels-Consonants Discrimination) was therefore included.

Domain of Visual Memory

Factorially, very little is known about the relationship between auditory and visual memory functions. Precise knowledge on this point would require a separate factor analysis; in the present study it was decided to include a few tests of visual memory which had previously been shown to be relatively independent of content (15). Such tests might be considered to typify more the visual process involved in memory rather than the material memorized. It was hoped that even such a few tests, if they were of the sort described, might be the first steps towards the solution of the problem of the relationship between auditory and visual memory functions and the establishment of a framework which could exist as a guide for future work. These tests were: Visual Memory Span (Test 28), already mentioned, Memory for Geometrical Drawings (Test 29), Memory for Boys' Faces (Test 30), and Memory for Limericks (Test 31). Because of its social implications, the correlation between Memory for Boys' Faces (Test 30) and Memory for Male Voices(Test 25) was awaited with particular interest.

Finally, the variables of Age (Test 32) and Intelligence Quotient (Test 33) were included to make some determination of the effects of biological growth on the response of the auditory mechanism and mental growth on the more complicated social responses in the auditory environment.

The Final Problem

The final nature of the complete problem might then be summed up as follows:

1. Is there a general auditory factor?
2. If not, is there any broad group factor which might be considered a practical approximation to a general factor?
3. Are there also, or instead, four functionally distinct group factors analogous to the physical concepts frequency, intensity, complexity, and duration?
4. Within each of these factorial categories, can it be shown that various other forms of the typical function of that category are

sufficiently saturated with the structural content of that function to warrant consideration only as sub-types of that category?

5. In addition to, or instead of, these four categories can it be demonstrated that complex auditory behavior involves functionally distinct auditory abilities corresponding to the postulated factors of rhythm, auditory analysis, auditory synthesis, and auditory memory?
6. To what extent do these four categories of pitch, loudness, quality, and time underly the complex auditory behavior required in social situations?
7. What can be shown factorially of the relation between auditory memory functions and visual memory functions?
8. What part do age and intelligence play in auditory function?

II. The Experiment

All the tests were group auditory tests and were given to members of both sexes in Whiting High School, Indiana, during two weeks of testing in September 1941. The subjects were volunteers who were promised a report on the state of their hearing. The acoustic situation was a school-room in a quiet part of the building. All the auditory tests were on records; the visual tests were given on a screen with a projector. About 25 subjects were tested at a time for sessions of 40 minutes per day.

THE TESTS

In setting up the problem the auditory field of this study was divided up into a number of domains on the basis of pre-existing evidence and a priori judgment. Each test therefore was chosen by this double criterion so as to have maximal representation in its particular domain.

I. The Pitch Domain:

Tests 1-4 called for judgments as to which of two stimuli in each item was the higher in pitch.

Test 1: *Pitch Discrimination for Pure Tones*

This was the pitch test in the Seashore *Tests of Musical Talent*, Series A (10). The stimuli were constant in intensity, complexity, and duration but varied in frequency. The score was the number of correct responses.

Test 2: *Pitch Discrimination for Complex Sounds*

The stimuli were complex sounds caused by setting in vibratory motion objects of differing resonance. The correct judgment for each item was taken to be the preference of the majority of subjects where a 2:1 judgment was obtained of one or other stimulus in each item. None of the physical factors were constant.

Test 3: Pitch Discrimination for Pure Tones of Short-Impulse

In each item two pure tones differing supra-liminally in pitch were compared. The stimuli varied from a duration beneath the duration threshold for pitch perception to one well above it. In each item intensity, complexity, and duration were held constant.

Test 4: Pitch Discrimination for Vocal Sounds

The stimuli in each item consisted of two monosyllabic vocal sounds vocalized on the same fundamental frequency. The correct response was determined on the basis of accepted standards of pitch differentials of vocal sounds (**2**).

II. Loudness Domain:

The instructions for these four tests (Tests 5-8) requested judgments as to which of the two stimuli in each item was the louder.

Test 5: Loudness Discrimination for Pure Tones

This was the loudness test in the Seashore *Tests of Musical Talent,* Series A (**10**). The stimuli were constant in frequency, complexity, and duration but varied in intensity. The score was the number of correct responses.

Test 6: Loudness Discrimination for Complex Sounds

This was the same test as Test 2 with the same scoring criterion applied to loudness judgments.

Test 7: Loudness Discrimination for Pure Tones of Short-Impulse

In each item two pure tones differing supra-liminally in loudness were compared. The stimuli varied from a duration beneath the duration threshold for loudness perception to one well above it. In each item frequency, complexity, and duration were constant.

Test 8: The Pitch-Loudness Function

In each item two pure tones of constant intensity, complexity, and duration but differing frequency were compared in loudness. The two frequencies for each item were so chosen that they would normally be heard as differing in loudness on account of the differential sensitivity to frequency (**3, 12**). A correct response would be one following the normal reaction.

III. The Time Domain:

This conventionally includes both filled and unfilled time.

Test 9: Sense of Time for Sound-filled Intervals

This was the Time test in the Seashore *Tests of Musical Talent,* Series A (**10**). In each item two tones of constant frequency, intensity, and complexity but differing duration are compared as to length. The score was the number of correct responses.

Test 10: Sense of Time for Intervals of Silence

This was the Time test in the earlier form of the Seashore *Tests of Musical Talent* (**9**). In each item the subject heard three clicks of constant frequency, complexity, and intensity but differing in tem-

poral arrangement. A correct judgment required deciding whether the silent interval between the first and second clicks was longer or shorter than the interval between the second and third clicks.

IV. The Rhythm Domain:

This includes motor rhythm and musical rhythm.

Test 11: *Motor Rhythm*

This was the Rhythm test in the Seashore *Tests of Musical Talent*, Series A (10). Each item called for a "same-different" judgment of two rhythmic patterns in which the stimuli are tappings of constant frequency, intensity, and complexity. The score was the number correct.

Test 12: *Musical Rhythm*

The subject decided whether short musical selections were being played in 2, 3, 4, or 6 time. The score was the number of correct judgments.

V. The Quality Domain:

This factor appeared to be represented also by several tests in other domains.

Test 13: *Quality Discrimination for Complex Tones*

This was the Timbre test in the Seashore *Tests of Musical Talent*, Series A (10). Each item called for a "same-different" judgment of two complex tones of the same fundamental frequency but differing weights of upper partials. Total intensity and duration of the stimuli remained constant.

VI. The Domain of Auditory Analysis:

Tests in this domain involved the ability to hear under different masking conditions.

Test 14: *Sound Breakdown*

For each item the subject judged how many of five speakers had spoken a word simultaneously. The score was the number correct.

Test 15: *Pure Tone Masking*

Two tones of widely differing frequency were sounded together. The upper tone was the more intense at first; the lower tone gradually became louder until it was audible together with the higher tone. The score obtained was the number correctly judged as having two tones.

Test 16: *Sensory Masking*

The subject was required to write down words heard against an increasingly loud buzzing background. The score obtained was the number of words heard correctly.

Test 17: *Intellective Masking*

The subject was required to write down isolated words heard against an increasingly loud background of a second continuous speaker. The score obtained was the number of words heard correctly.

Test 18: Auditory Discrimination for Vowels and Consonants

This was the Wepman *Test of Auditory Discrimination* used in the Speech clinic at Billings Hospital, University of Chicago. In each item the subject made a "same-different" judgment for two words differing in a vowel or consonant. The score obtained was the number of correct responses.

VII. The Domain of Auditory Synthesis:

These tests tapped the ability of the organism to resist distortion of meaning due to disturbance of the temporal sequence of sounds.

Test 19: Rapid Spelling

The subject was required to write down familiar words which had been spelled out very rapidly. The score obtained was the number of words correctly understood.

Test 20: Singing

The subject wrote down the words of a short vocal selection sung with piano accompaniment. The score obtained was the number of words correctly understood.

Test 21: Haphazard Speech

The subject wrote down the words of a short phrase spoken with unusual inflection and pitch changes. The score obtained was the number of words understood.

Test 22: Illogical Grouping

The subject was required to write down the words of a short phrase spoken with a grouping arrangement contrary to the sense of the passage. The score was the number of words understood.

VIII. The Span Domain:

This included both auditory and visual memory span.

Test 23: Auditory Fusion Memory Span

The auditory stimuli were nonsense-syllables of increasing length such that the words were all vocalizable. The letters were presented one per second. The score obtained was the number of words correctly written down.

Test 28: Visual Fusion Memory Span

The visual stimuli were nonsense-syllables of increasing length presented a letter at a time on a projector screen. Each letter was in view for about half a second. The words were all vocalizable. The score obtained was the number of words correctly written down.

IX. The Auditory Memory Domain:

This involved laboratory and social tests of memory.

Test 24: Memory for Female Voices

A number of female speakers read excerpts from homogeneous scientific material in random order of vocal re-appearance. For each

speaker the subject decided whether or not he had heard that speaker previously in the test. The score was the number correct.

Test 25: *Memory for Male Voices*

This was the same test as the previous one except that male voices were used.

Test 26: *Memory for Pitch Gestalt*

This was the Tonal Memory test in the Seashore *Tests of Musical Talent,* Series A **(10)**. Each item required a comparison of two short melodic phrases which were identical except that one tone in one of the phrases was supra-liminally changed in pitch. The subject was required to name the ordinal number of the note changed. The score was the number correct.

Test 27: *Memory for Emphasis*

The subject heard a two-minute extract read with certain words markedly emphasized. He was required to identify these words on a written script at the conclusion of the reading. There were three such extracts. The score was the number of words correctly identified minus those incorrectly identified.

X. The Visual Memory Domain:

The memory tests chosen had previously been shown to be relatively independent of content.

Test 29: *Memory for Drawings*

The subject is shown a number of geometrical drawings on a screen and is later required to identify these drawings among similar drawings. The score was the number correctly identified.

Test 30: *Memory for Boys' Faces*

This was the same as the previous test with boys' faces instead of drawings.

Test 31: *Memory for Limericks*

The subject was shown a number of limericks on the screen and was subsequently required to write in the last line of each limerick on the response sheet. The score was the number of last lines completely correctly written in.

XI. Miscellaneous:

Test 32: *Age*

The ages of the subjects ranged from 15 to 19 years.

Test 33: *Intelligence Quotient*

An I.Q. for each subject based on an Otis or Henmon-Nelson test was available from the school records.

III. The Factor Analysis

All tables of results are given at the end of the article. Each test was so scored that high score indicated high ability. The Pearson

product-moment correlations are shown in Table 1. It is seen that the auditory field is virtually a positive manifold. Inspection of the correlations led to the rejection of Test 15, Test 24, Test 30, and Test 32 from the subsequent analysis. It is of interest to note that the correlation between Memory for Male Voices (Test 25) and Memory for Boys' Faces (Test 30) is insignificant.

The correlational table was factored by the grouping method recently developed in the Thurstone computing laboratory (14). This method is based on the same principles as the centroid method but the factors extracted are nearer the final rotated meaningful primary factors. It is possible to rotate from one set of results to the other by orthogonal transformations. Nine factors were extracted and rotated by oblique rotations with unextended vectors in accordance with the demands of simple structure. Eight unambiguous interpretable planes emerged with the ninth factor a positive unidentifiable plane. The loadings on the centroid factors are given in Table 2; the rotated factor matrix is presented in Table 3. In Table 4 is shown the transformation matrix (Λ) leading from the centroid matrix (F_c) to the final rotated matrix (V) by the function

$$V = F_c \Lambda .$$

The correlations between the primary vectors obtained by the equation

$$R_{pq} = D(\Lambda'\Lambda)^{-1} D'$$

are shown in Table 5, where D is a diagonal matrix such that the diagonal entries in R must be unity.

IV. Interpretation of Factors

Each factor will be discussed in turn. For the most part loadings below .30 will not be considered.

Factor A

Test		
4.	Vocal Pitch Discrimination	.70
3.	Short-Impulse Pitch Discrimination	.67
1.	Pure Tone Pitch Discrimination	.67
13.	Quality Discrimination	.45
26.	Tonal Memory	.42
2.	Complex Tone Pitch Discrimination	.28

In every one of these tests the two stimuli in each discriminatory judgment differ in the frequencies of the component parts of the tones. In Tests 3, 1, and 26 the difference is in the frequency of the fundamental and only frequency. In Test 4, 13, and 2 the difference

is either in the frequencies of the fundamental or the overtone structure. It would appear, therefore, that pitch judgments are some function of all audible frequencies in a stimulus. This functional unity is therefore termed a *pitch-quality* factor. In other words, the pitch of a sound is the weighted impression of both the fundamental and overtone frequencies. Through training the pitch can be fixed as that of the fundamental; through training the effect of the overtone frequencies can be restricted to what are conventionally called timbre or quality judgments. Basically, however, the same functional system subserves both applications of the auditory frequency apparatus. The pitch and quality division is the result of a physically-derived view; psychologically, the division is probably an artifact.

This factor will be known in this study as the *PQ* (pitch-quality) factor.

Factor B

Test		
7.	Short-Impulse Loudness Discrimination	.48
8.	Pitch-Loudness Function	.47
6.	Complex Sound Loudness Discrimination	.45
5.	Pure Tone Loudness Discrimination	.42
25.	Memory for Male Voices	.40
33.	Intelligence Quotient	.31

The interpretation that most readily suggests itself is a *loudness factor*. In Tests 7 and 5 it can be shown that the tones of each item differ in intensity; it is of significance that the same functional system that determines successful loudness discriminations in these two tests also turns out to be the tendency among the majority in a random population to decide between the two stimuli in Test 6 which cannot easily be measured in physical terms. Of further anatomical interest is the finding that success in the other loudness tests is bound up with the normal differential reaction to frequency in Test 8. The same system which enables the subject to discriminate the different intensities of a single frequency also enables him to discriminate the loudnesses of two different frequencies of the same intensity.

It is seen that neither time nor complexity is particularly important for loudness judgments. It appears that the crux of the loudness function is the average strength of the psychological response to a given frequency. *Memory for Male Voices* would apparently depend upon the loudness level and loudness inflections characteristic of the individual speaker. The significance of the loading of the intelligence quotient is taken to be an indication of the possible perceptual level of the loudness function as opposed to the sensory level conventionally assumed.

This factor will be known as the *L* (loudness) factor in this study.

Factor C

Test		
10.	Unfilled Time	.50
5.	Pure Tone Loudness Discrimination	.48
6.	Complex Sound Loudness Discrimination	.38
9.	Filled Time	.38
14.	Sound Breakdown	.32

This does not appear to be a time factor but is the closest approximation to a time factor which is supported by the factorial evidence. The essential element common to all the test-projections is a mass quantity dependent on occurrence in time for its formation. The physical analogy of the integral symbol, $\int$, is strongly suggested in the interpretation of this factor. The factor is therefore called the *Auditory Integral for Perceptual Mass* factor.

The Auditory Integral for Perceptual Mass factor is defined operationally in the following way: Consider any auditory event occurring over a short period of time as being known by its average instantaneous loudness parameter and a time parameter. The primitive quantitative mass outlined by these two parameters becomes known in consciousness by an integrative process of the auditory mechanism, an integration of the mass quantity between the limits of the beginning and end of the auditory stimulation. If the parameters of two auditory events are identical, then the magnitude of the integral is the same for the two events. If the average instantaneous loudness parameters of two auditory events are identical but the time limits are narrower for one event, the integral for the shorter stimulus will be smaller in magnitude. If the time limits are the same for auditory events but the average instantaneous loudness parameter is smaller for one event, the integral will be smaller for that event.

The mass quantity being integrated by this factor or ability appears to consist both of the loudness response corresponding to the intensity of a sound and the positive after-image of that energy stimulation. In *Unfilled Time* (Test 10) three clicks form two intervals

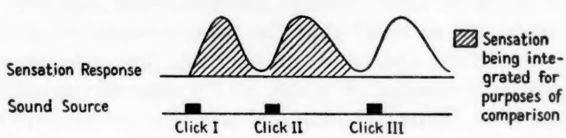

Figure 1

of physical silence; psychologically, the positive after-image of a click lingers after cessation of the physical sound source; comparison of the integral for the after-image between the first and second clicks with that between the second and third clicks would yield a larger integral for the more widely separated clicks (see Figure 1).

In *Filled Time* (Test 9) two sounds have the same instantaneous loudness, but due to the fact that one of the sounds lasts longer, when the ear integrates the total perceptual mass between the two differing limits the integral will be larger for the sound with wider limits. In *Pure Tone Loudness Discrimination* (Test 5) two sounds have the same duration or the same limits for integration, but, since the average instantaneous loudness differs, the integral will be larger for the louder sound. In *Complex Sound Loudness Discrimination* (Test 6) there is presented a more variable picture of the process described above for *Pure Tone Loudness Discrimination.* The complexity of the sounds with the intensity rise-and-fall effects characteristic of the striking of gongs and similar media of vibration makes the integration over approximately equal limits a more difficult matter. The saturation of this test with the Integral factor is correspondingly significant but smaller than that of the *Pure Tone Loudness Test.*

The interpretation of this factor is materially facilitated by a consideration of the behavior of the four loudness tests in this domain. The loadings are:

Test		
5.	Pure Tone Loudness Discrimination	.48
6.	Complex Sound Loudness Discrimination	.38
7.	Short-Impulse Loudness Discrimination	.21
8.	Pitch-Loudness Function	.08

From the foregoing it would follow that the Integral factor is most effective in loudness discrimination with pure tones, less so with complex sounds, barely effective with short tones, and not at all operative with the tones of the pitch-loudness function. It has already been shown that pure tone loudness discrimination involves this factor fairly considerably and that complex sound loudness discrimination would involve it to a lesser extent. In the case of short impulse loudness discrimination the two tones have identical time limits for integration but differ in their average instantaneous loudness; however, the time limits are so short that the mass-integrative ability of the ear is afforded little opportunity to operate. The loading of this test on this factor is consequently small. In the case of the pitch-loudness function the time limits are identical and the intensity parameter is the same in that the physical energies of the two sounds are constant;

the integrative power of the ear is therefore not given any opportunity at all to function. The loading of this test on this factor is, as would be expected, negligible.

The nature of the psychological excitation represented pictorially as the curves in Figure 1 above is at present indefinite. It seems possible that this type of response is similar to the corresponding response postulated in the visual field. Current evidence would place the onset of the auditory sensation at its maximum intensity between .12 sec. to .50 sec. after the onset of the physical stimulus; termination of the auditory sensation occurs about .14 sec. after the physical stimulus is removed. These time values depend mainly on the intensity of the stimulus. It is to be supposed, therefore, that the integration of perceptual mass described above commences between .12 sec. and .50 sec. after the physical stimulus is first sounded. The memory image of the stimulus, however, persists beyond the duration of actual sensation, so that the integration is probably in terms of both the sensation and the memory image. The entire integral process is presumably of short duration, that is, less than a second, if integration is the basis for comparison of one stimulus with another.

It is of interest to note that this factor appears to offer factorial evidence of a relation previously postulated both in the auditory and visual fields. Insofar as the integral is affected with equal ease by the time and the intensity parameters, it is in line with the Bunson-Roscoe Law in vision and the Lifshitz Law in hearing, both of which postulate that:

$$It = K$$

where I is the intensity, t the time of the stimulus, and K is a constant.

Since these four loudness tests have about the same loadings on the L (loudness) factor and the correlation between this Auditory-Integral factor and the Loudness factor is insignificant, it may be stated that this Integral factor is in no way a form of the loudness function; the two are independent and the integrative function is operative on another level over and above the loudness level.

This factor will be known as the *AI* (Auditory Integral) factor in this study.

Factor D

Test		
21.	Haphazard Speech	.59
22.	Illogical Grouping	.58
20.	Singing	.56
17.	Intellective Masking	.30
16.	Sensory Masking	.22

This factor appears to underlie both the domains of auditory synthesis and analysis. Instead of one auditory ability enabling the organism to resist distortion of words due to temporal disarrangement, and another ability for resistance to masking noises obscuring meaning, there is apparently a more central ability which serves both purposes. This factor will be known as the *AR* (auditory resistance) factor. This functional system is probably widely operative in most auditory environments in social life.

Factor E

Test		
28.	Visual Fusion Memory Span	.53
19.	Rapid Spelling	.52
17.	Intellective Masking	.31

The common characteristics of these tests is the rapidity with which the stimuli have to be received in order to be perceived. In each case stimuli are presented rapidly, and if the subject is able to interpret the rapid sensations as well as being able to receive them at that speed in the first case, he is able to perform the task. The importance of this factor is partly the fact that it denotes an ability which transcends sense modality and operates equally well in Test 28, a visual test, and Test 19, an auditory test.

This factor will be known in this study as the *SC* or *Speed of Closure* factor.

Factor F

Test		
23.	Auditory Memory Span	.49
9.	Unfilled Time	.38
5.	Pure Tone Loudness Discrimination	.34
10.	Filled Time	.27
2.	Complex Pitch Discrimination	.26

This auditory factor apparently represents the actual mechanics by which the auditory memory span is formed; time, loudness, and pitch elements appear concerned here. This is not a memory factor primarily. This factor will be known as the *ASF* or Auditory Span Formation factor in this study.

Factor G

Test		
16.	Sensory Masking	.36
26.	Tonal Memory	.36
23.	Auditory Fusion Memory Span	.32
11.	Motor Rhythm	.29
17.	Intellective Masking	.25
28.	Visual Fusion Memory Span	.22

The interpretation offered here is that this factor is the true memory element in the *Memory Span* factor. All the projected tests on this factor involve immediate recall of material within the span. The ability appears to hold for a variety of auditory stimuli and to extend to stimuli of the visual domain. This is probably a general span factor independent of sense modality and has no relation to any other factor in the analysis.

This factor will be known as the *GS* (General Span) factor.

Factor H

Test		
16.	Sensory Masking	.56
31.	Memory for Limericks	.51
33.	Intelligence Quotient	.49
29.	Memory for Drawings	.34

In Test 16 the material is presented auditorily; in Tests 31, 33, and 29 the stimuli are visual. In each of these tests the subject is presented with an extensive array of possible stimuli and is required to give the greater part of his attention to a selected few of these stimuli. The comprehension of the crucial stimuli depends on a kind of mental alertness and ability to consider incidental stimuli only as means of obtaining the crucial stimuli. This ability enables the subject to reproduce the crucial stimuli at a later point, either immediate or delayed, given certain partial clues from the incidental stimuli immediately preceding the crucial stimuli.

This factor appears to be a closure effect transcending sense modality, dependent on partial clues from the source of stimulation. The factor will be known as the *IC* or *Incidental Closure* factor.

Factor J

This appeared to be a positive uninterpretable plane.

V. *Evaluation of Hypotheses*

1. *Problem*: "Is there a general auditory factor, or any broad group factor which might approximate the general factor?"

There is no support for either a first- or second-order general auditory factor. No factor has loadings on tests representative of the various primary factors. The indications are that auditory function occurs on different levels and that any single factor will be a poor approximation to a complete picture of hearing.

2. *Problem*: "Are there four functionally distinct group factors analogous to the physical concepts frequency, intensity, complexity, and duration of sound waves?"

None of these factors are verified as being simple sensory auditory factors. The pitch and quality tests are subsumed under a single

functional system, pitch-quality. The loudness factor is probably more of a perceptual process than has been suspected and is certainly more complex than previously supposed. The loudness function involves at least two functional unities, the average strength of the differential psychological response to sound in the *L* factor, and the integration of energy disturbance over given time limits in the *AI* factor. The pitch function, in contrast, is factorially quite simple. The time tests do not specifically define any time factor but play integral parts in the *AI* factor.

4. *Problem*: "What is the relation between different aspects of each of the foregoing four categories?"

The four tests of pitch used are apparently legitimate forms of the same basic functional system and the same ability determines discriminatory judgments in all of them. Quality discrimination turns out to be another form of the same system as underlies pitch manifestations. The four tests of loudness are equally good indications of the working of the loudness function; it seems, however, that in making loudness judgments in certain situations an integration of a more primitive perceptual mass is involved which is not involved in other loudness situations. The two tests of time behave in the same manner in the analysis, that is, ancillary to other functions.

5. *Problem*: "Does complex auditory behavior involve rhythm, auditory analysis, auditory synthesis, and auditory memory as distinct functional auditory abilities?"

No rhythm factor is defined; the inference would seem to be that the rhythmic sense is kinaesthetic or central in nature. Both the postulated abilities of auditory analysis and auditory synthesis break down to a single entity, the *AR* factor. Auditory memory is important, yielding several memory factors even with comparatively few memory tests.

6. *Problem*: "To what extent do the simpler tests of pitch, quality, loudness, and time underlie more complex social auditory behavior?"

These tests are involved principally only in the *PQ*, *L*, and *AI* functions. These latter functions are not related to the other auditory functions which appear to be more directly concerned with social audition. This finding is cause for concern regarding current practice of assuming that complex auditory behavior is explainable in terms of performance on the simpler functions.

7. *Problem*: "What is shown here regarding the relation of visual and auditory memory functions?"

Three of the four tests involving memory tests were common to both the visual and auditory field. Both kinds of memory function tend to be specific, but this specificity is largely independent of sense modality.

8. *Problem*: "What part do age and intelligence play in auditory function?"

Biological growth in this age range had no effect on goodness of performance of auditory mechanisms. The effect of intelligence was less marked in the complex auditory functions than might have been expected; the relation of intelligence to the loudness factor was rather surprising.

VI. *Summary of the Investigation*

This investigation was concerned with the factorial description of correlated individual differences in the auditory field. Several hypotheses as to the nature of auditory function among normally-hearing subjects were being empirically investigated. A population of 200 high-school individuals were given 27 group auditory tests and 4 group visual memory tests; the age and intelligence quotient of each subject was also considered. The correlations of 29 of these tests were obtained and factored to nine factors. The configuration was rotated to a meaningful oblique simple structure. Interpretation of eight factors is offered.

A large number of conclusions might be drawn from this initial factorial study of the auditory field. Some of the main conclusions drawn are mentioned below:

1. There is no general auditory factor; no group factor appears to offer a practical approximation to a general factor.
2. The pitch and quality categories break down to a single basic process; the loudness tests appear more complex than the pitch-quality tests and require at least two distinct processes for successful performance; the time tests do not define any time factor but play an integral part in several of the other auditory factors.
3. Rhythm does not appear to be an auditory factor primarily; adequate auditory response-patterns in complex auditory situations do not appear to require rhythm to any appreciable extent.
4. The phenomena of auditory analysis and auditory synthesis are subsumed under a single functional system, a resistance to distortion factor.
5. Both auditory and visual memory functions appear highly specific but overlap in their specificity to produce several central fac-

tors independent of sense modality, notably a general span factor.

6. For high-school subjects neither age nor intelligence play any important part in most of the auditory functions.
7. The conventional auditory acuity tests have little predictive value for auditory behavior in more complex social situations for normally-hearing subjects.

REFERENCES

1. Fletcher, H. *Speech and Hearing.* New York: D. Van Nostrand Co., 1929.
2. ———. *Ibid.*, p. 28.
3. ———. *Ibid.*, p. 141.
4. Fowler, E. P. *Medicine of the Ear.* New York: Thos. Nelson and Sons, 1939.
5. Hartmann, G. W. *Gestalt Psychology.* New York: Ronald Press Co., 1935. Chap. 7.
6. Karlin, J. E. Music Ability. *Psychometrika*, 1941, 1: 61-65.
7. Seashore, C. E. *Psychology of Music.* New York: McGraw-Hill, 1938.
8. ———. *Ibid.*, p. 138.
9. Seashore, C. E. *Manual for Measures of Musical Talent.* Chicago: C. H. Stoelting Co.
10. Seashore, C. E., Lewis, D., Saetveit, J. G. *Manual for The Seashore Measures of Musical Talents.* New Jersey: RCA Manufacturing Co., 1939.
11. Stevens, S. S. and Davis, H. *Hearing.* New York: McGraw-Hill, 1938.
12. ———. *Ibid.*, p. 124.
13. ———. *Ibid.*, p. 98.
14. Thurstone, L. L. Grouping Method of Factoring. Unpublished.
15. ———. Memory Factor Study. Unpublished.

Addendum

Grateful acknowledgment is made of the constant advice and supervision of Dr. L. L. Thurstone, the assistance in construction of the phonograph records by Mr. Ernest A. Ewers of the Chicago Bell Telephone Engineering Department, and the aid given by Mr. Ledyard R. Tucker in the statistical analysis. The investigation was possible through the generous financial Aid of the Carnegie Foundation for research funds, made available through Professor Thurstone, for making phonograph records and other equipment as well as for the cost of tabulating machine services in the factorial work. Acknowledgment is also made to the Psychometric Laboratory and the Social Science Research Committee of the University of Chicago for the use of equipment and for computing.

This report is an abbreviated account of selected portions of a Ph.D. thesis in the Department of Psychology at the University of Chicago. There appears to be a rather general interest in the results of this study both on the part of psychologists and investigators in other fields; a complete account of the investigation will therefore appear in monograph form.

TABLE 1
Correlation Matrix*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1																	
2	237																
3	628	361															
4	575	313	686														
5	242	195	275	256													
6	035	015	–026	034	358												
7	008	264	156	112	344	299											
8	115	008	108	098	248	183	291										
9	072	126	174	009	309	–025	155	079									
10	344	196	283	288	462	238	254	148	276								
11	213	131	241	197	222	109	119	010	002	230							
12	218	103	286	266	157	014	108	019	035	035	095						
13	337	130	302	299	151	009	046	–069	138	153	155	099					
14	156	199	180	134	188	063	199	014	159	196	061	086	254				
15	004	071	124	168	004	–009	–005	–047	101	–068	060	074	138	–116			
16	106	110	142	136	108	038	058	014	–118	048	202	039	045	195	–080		
17	153	054	236	126	130	133	159	078	004	084	208	068	148	203	063	353	
18	017	114	014	039	183	124	051	040	123	213	039	–005	040	181	125	086	190
19	206	156	263	312	079	–005	010	075	–082	081	192	132	061	141	–005	232	314
20	069	218	166	169	018	–049	140	–038	025	059	279	–013	059	144	072	241	264
21	–002	044	088	106	172	076	037	031	060	147	171	107	–013	140	116	167	205
22	142	184	241	221	299	150	199	–004	093	236	276	–056	204	316	–116	379	309
23	138	216	227	122	136	–057	062	–050	186	134	255	146	102	061	004	250	186
24	074	051	101	–014	047	037	095	061	014	057	102	–086	058	007	009	000	007
25	118	182	091	103	131	133	211	247	010	113	192	020	088	118	–125	023	090
26	410	184	448	455	235	080	132	054	152	234	396	132	341	258	030	313	238
27	–019	207	021	–011	066	100	227	033	198	251	102	–002	008	243	–173	236	142
28	201	085	253	258	171	037	175	131	077	088	137	084	120	066	037	212	309
29	192	084	241	229	211	180	105	052	073	185	153	064	118	144	102	271	176
30	134	025	082	089	133	153	044	092	041	110	–004	138	099	052	–039	124	106
31	192	099	222	202	085	137	228	134	–003	072	113	143	–048	119	–020	348	176
32	020	–001	–042	018	–008	116	–010	–006	125	109	101	–095	072	042	069	056	062
33	147	085	213	294	216	209	247	212	–138	168	190	245	047	069	032	441	266

* The decimal points preceding all correlation coefficients have been omitted.

TABLE 1
Correlation Matrix (continued)*

	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
18																
19	098															
20	188	225														
21	040	137	348													
22	176	163	400	462												
23	109	156	193	057	199											
24	069	014	110	023	021	–006										
25	068	149	199	112	095	163	367									
26	107	258	282	197	364	178	065	237								
27	074	152	169	061	231	216	014	054	163							
28	058	421	121	–085	080	358	–039	141	172	138						
29	040	180	092	204	202	042	019	004	187	162	198					
30	070	059	054	171	101	027	–101	092	096	077	044	167				
31	067	313	086	009	174	180	005	072	152	175	167	326	085			
32	023	–088	027	056	153	–008	–003	–004	261	056	–069	–132	016	–196		
33	093	344	009	064	194	129	–005	095	174	100	170	298	089	394	–236	

* The decimal points preceding all correlation coefficients have been omitted.

TABLE 2
Unrotated Factorial Matrix

	I	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*	h^2
1	71	–21	–11	00	–12	–09	–17	–09	07	61
2	41	06	02	–13	07	–07	22	17	11	27
3	78	–11	–18	–05	–07	–03	16	09	06	68
4	77	–16	–21	07	–26	01	10	08	–02	75
5	47	02	53	–07	–01	–10	05	–08	–22	53
6	14	08	50	17	–20	03	–08	–11	–07	36
7	28	14	47	–02	01	09	11	23	18	40
8	16	–06	39	12	–07	13	–05	16	17	25
9	23	01	25	–36	27	–10	19	–11	11	37
10	46	04	37	–13	01	–14	10	–16	–10	43
11	39	23	05	–02	–01	05	–22	17	–17	28
12	32	–12	–05	12	–01	–04	12	17	–08	18
13	41	02	–13	–18	–03	–10	–17	–23	08	28
14	33	30	07	–08	02	–12	09	–21	21	28
16	26	49	–10	44	17	–22	–11	06	–08	61
17	33	39	03	14	10	22	–13	–08	–08	36
18	15	20	16	–09	02	13	07	–13	–16	14
19	42	18	–18	30	07	42	11	–05	04	53
20	24	49	–16	–17	–01	18	05	21	–05	43
21	16	48	02	–13	–26	06	15	08	–22	38
22	36	63	06	–09	–11	–10	04	–04	–12	57
23	34	16	–08	–05	50	–01	–05	19	–19	44
25	24	10	21	–07	–04	20	–21	30	19	29
26	58	27	–10	–09	–05	–12	–22	04	09	50
27	17	31	18	02	29	–05	17	–04	16	27
28	40	01	02	19	35	39	–13	–07	03	49
29	33	19	08	28	01	–07	10	–12	06	26
31	28	17	07	47	10	–03	17	10	16	38
33	37	13	18	60	–08	–06	03	15	–12	58

TABLE 3
Rotated Factorial Matrix

Centroid Factor	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*
Psychological Factor	*PQ*	*L*	*AI*	*AR*	*SC*	*ASF*	*GS*	*IC*	*J*
1	69	09	00	–08	02	01	17	–02	–03
2	28	05	04	08	–07	26	–07	10	24
3	67	04	–03	05	03	16	03	03	04
4	70	07	–07	08	04	04	–09	03	–10
5	15	42	48	–03	–03	34	–02	00	01
6	–07	45	38	–01	02	–07	–02	07	03
7	–01	48	21	05	–03	18	–07	11	40
8	–01	47	09	–12	02	03	–03	05	26
9	13	01	38	–07	–01	38	–04	–08	29
10	24	23	50	00	–03	27	–03	02	07
11	10	22	01	24	02	11	29	–01	–06
12	22	07	–09	–06	–04	18	–07	11	–05
13	44	–09	16	06	03	–05	22	–08	00
14	22	–05	–32	19	01	–03	06	18	27
16	–07	–06	–02	22	–06	–01	36	56	03
17	–06	06	15	30	31	–02	25	07	00
18	–08	06	28	20	17	10	–03	–09	–05
19	08	–05	–04	20	52	–06	–02	09	02
20	–02	–04	–09	56	09	04	12	–02	08
21	–04	05	12	59	–03	–04	–09	01	–07
22	06	03	28	58	–05	–01	16	19	06
23	–03	–08	–03	05	10	49	32	06	00
25	04	40	–08	10	03	01	20	–07	28
26	42	07	05	26	–07	–02	36	10	12
27	–09	–01	25	08	06	20	05	26	34
28	01	06	06	–08	53	16	22	–05	04
29	12	03	20	06	08	–02	01	34	10
31	01	09	00	–02	06	02	–02	51	23
33	02	31	03	–01	00	01	03	49	–03

TABLE 4
Direction Cosines of the Reference Vectors

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*
I	68	18	20	08	14	25	15	07	06
II	–35	–09	17	80	01	–19	27	33	21
III	–31	77	62	–22	–08	25	–10	02	31
IV	–20	09	–16	–26	15	–26	–01	67	–10
V	–33	–29	06	–37	26	66	35	11	19
VI	–31	14	–11	21	85	–13	–16	–54	–07
VII	–02	–29	18	10	04	30	–86	23	15
VIII	–13	41	–68	09	–41	31	06	12	21
IX	26	02	–12	–18	–04	–37	02	25	86

TABLE 5
Correlations Between the Primary Vectors

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*	
A	1.00									*PQ*
B	20	1.00								*L*
C	–08	–16	1.00							*AI*
D	18	13	–07	1.00						*AR*
E	34	28	–20	08	1.00					*SC*
F	20	06	–18	32	18	1.00				*ASF*
G	03	–07	16	–01	03	01	1.00			*GS*
H	19	24	–07	11	41	13	–02	1.00		*IC*
J	–08	–26	–04	02	–03	–02	00	–33	1.00	*J*

PSYCHOMETRIKA—VOL. 7, NO. 4
DECEMBER, 1942

TEST SCORES EXAMINED WITH THE LEXIS RATIO

HAROLD A. EDGERTON AND KENNETH F. THOMSON
THE OHIO STATE UNIVERSITY

The Lexis Ratio is discussed in its application to distributions of test scores where the items of the test can be assumed to be of equal difficulty. The ratio indicates the extent to which inter-individual variation operates as a source of the variance. The concept is related to the Lexis, Bernoulli, and Poisson distributions and illustrated by urn schemata. The Ratio is applied to the scores of 560 university freshmen on the *Robinson Reading Test*. The relation of the Lexis Ratio to the Kuder-Richardson estimation of reliability is also discussed and the latter authors' case IV is rewritten explicitly in terms of the Ratio.

The Lexis ratio is a statistic used to show whether a distribution of observations has hypernormal, normal, or subnormal dispersion. The hypernormal or Lexis distribution is of particular interest in connection with test scores, since with this type of distribution one may infer that differences among the individuals tested are present as a source of variance. A hypernormal (Lexis) dispersion is obtained when the probability of occurrence of an event is constant from trial to trial within a set, but varies from set to set. Put into terms of test scores, the term *trial* refers to a test item, the term *set* to the individual taking the test.

A normal (Bernoulli) dispersion may be said to reflect no real individual differences, since normal dispersion is to be expected when the probability of occurrence of an event not only is constant from trial to trial within a set, but also is constant from set to set. In terms of testing, this would correspond to the responses of a population of identical individuals on a test all items of which are of equal difficulty.

A subnormal (Poisson) dispersion does not enter the picture of testing, since it is obtained when the probability of the occurrence of an event varies from trial to trial, but the several probabilities of every one set of trials are identical with those of the corresponding trials of every other set.

In order to apply the Lexis ratio technique, the test score for the individual must be expressed as the per cent or proportion of "successful" trials. The items are, by assumption, of equal difficulty.

Urn schemata can be used to illustrate the Lexis distribution.

Let us suppose that five urns have been filled with black and white balls, and so maintained that the probabilities of drawing a white ball from the first urn will be 1/10, from the second 2/5, from the third 1/2, from the fourth 4/5, and from the fifth 7/10. From each urn a set of 10 drawings of one ball each will be made.

The score for a *set* is the ratio of white balls drawn to the total of balls drawn. Within each set (urn), the probabilities of a white ball are constant from trial to trial. But the probabilities of the drawing of a white ball vary from set to set. The scores for the five sets will differ, and the dispersion will be hypernormal.

Similarly, in a test where the items are of equal difficulty and where the individuals taking the test differ in ability, one may expect a Lexis or hypernormal dispersion of test scores.

Thus the Lexis ratio would seem to afford a convenient means for the examination of percentage test scores of various samples. A limitation of convenience in the use of this technique is that the test must permit a count of "items attempted." Where items are so arranged or instructions so worded that the testee may skip an item or items without penalty, the increased clerical work necessary to determine "items attempted" might weigh against the use of the technique on economic grounds.

The Lexis ratio,

$$L = \sigma/\sigma_B , \tag{1}$$

compares the obtained dispersion, σ, of percentage scores on a test with the theoretically expected dispersion, σ_B, calculated from the mean percentage success and the mean number of items attempted.

When $L > 1$, we have a Lexis distribution and may infer that a portion of the variance is due to differences among the individuals. When $L = 1$, we have a Bernoulli distribution and may infer that the observations differ from their mean value only because of chance factors.

The formula for the Lexis ratio may also be written

$$L = \frac{\sigma'}{\sigma'_B}, \tag{2}$$

where σ' is the observed standard deviation of a series of scores, each score expressed as a proportion or per cent. This value may be calculated by any of the usual procedures. The value of σ'_B is given by the formula

$$\sigma'_B = \sqrt{\frac{pq}{s}}, \tag{3}$$

where p = mean value of scores, (each one expressed as proportion or per cent),

$q = (1 - p)$ for scores expressed as proportions, or $(100 - p)$ if the scores are expressed as per cents,

s = number of items attempted by each individual, and is the denominator used in obtaining proportion or per cent scores. In case all individuals did not attempt the same number of items, s may be taken here as the mean number of items attempted.

As a demonstration of the application of the method to actual data, the technique was used on the scores in "percentage comprehension" on the *Robinson Reading Test.**

The *Robinson Reading Test* consists of a standard reading dealing with a certain subject and a separate set of questions covering factual material presented in the reading. At the end of ten minutes spent in reading the material, the testee is asked to indicate the number of lines of the material he has read in the period. This mark indicates his average reading speed for the given period. The reading material is then removed and the set of questions is given with the instructions that he is to answer or attempt to answer all the questions up to and including the end point reached during the reading period. A scale is included on the question sheet to indicate the lines of reading covered by each question. The comprehension score is calculated by dividing the number of items answered correctly by the number of items attempted, and hence is amenable to investigation by the Lexis ratio.

The sample used for this demonstration is made up from 560 *Robinson Reading Tests* given to College of Education freshmen at The Ohio State University in the Autumn quarter of 1941.† The first question to be answered here would be: Do the comprehension scores observed in this sample reflect individual differences? To answer the question, the entire 560 cases were utilized. From the data of the sample:

$$N = 560$$

$$p = 65.71 \text{ per cent (mean per cent correct answers)}$$

$$q = 100 - p = 34.29$$

* Robinson, F. P. and Hall, P., Studies of higher-level reading abilities, *J. Educ. Psychol.*, 1941, 32, 241-252.

† These data were made available through the courtesy of L. L. Love, Junior Dean of the College of Education.

$$s = 24.78 \text{ (mean number of items attempted)}$$

$$\sigma' = 13.691$$

$$\sigma'_B = \sqrt{\frac{(65.71)\,(34.29)}{24.78}} = 9.536$$

$$L = \frac{13.691}{9.536} = 1.436.$$

May we now conclude that this distribution of errors has hypernormal dispersion and that individual differences rather than item differences on the test are reflected? The probable error of the Lexis ratio is given by the formula:

$$P.E._L = \frac{.4769\,L}{\sqrt{N}}.$$

For the Lexis ratio above,

$$P.E._L = \frac{(.4769)\,(1.436)}{\sqrt{560}} = .0289.$$

The critical ratio sought is one which will answer the question: Is it within reason to assume that the observed Lexis ratio could have been drawn from a universe in which the true value of the Lexis ratio is 1.00? Hence we may take the function

$$C.R. = \frac{L-1}{P\,E_L} = \frac{.436}{.0289} = 15.1.$$

On the basis of such a critical ratio, it may be concluded that the Lexis ratio of 1.436 differs significantly from 1.00, and so it is inferred that the observed differences in reading comprehension may in part be ascribed to differences among the individuals.

It may be noted that in the *Robinson Reading Test,* the number of items attempted was a function of reading speed. The question then arose as to whether or not the comprehension score reflected individual differences when reading speed was held relatively constant. In order to investigate this aspect, eight sub-samples were selected from the total sample.

Groups B, D, and G are the more heterogeneous in terms of variation in reading speed. The evidence does suggest that the test does not reflect individual differences so well for slower readers as for average and rapid readers. Such evidence is only suggestive. The hypothesis might well be investigated in a more appropriate situation.

The results in detail for these eight sub-groups are shown in Table 1. In all cases L is greater than unity.

TABLE 1
The Lexis Ratios of the Subgroups

	A	B	C	D	E	F	G	H
No. of cases	24	56	37	87	42	80	40	17
No. of items attempted	31	30,31	28	24,25	24	22	18,19 20	18
Av. Reading Speed (W.P.M.)	260	268	225	204	199	173	154	144
s (mean no. of items attempted)	31	31.4	28	24.5	24	22	18.9	18
p	68.75	67.94	66.15	65.78	66.19	64.44	64.38	64.56
σ'	12.10	14.58	9.98	12.36	13.14	12.08	15.56	16.67
σ'_B	8.32	8.32	8.94	9.58	9.66	10.21	11.02	11.27
L	1.45	1.75	1.12	1.29	1.36	1.18	1.41	1.48
$P.E._L$	.141	.112	.088	.066	.100	.063	.106	.171
$(L-1)/(P.E._L)$	3.21	6.69	1.33	4.39	3.60	2.86	3.87	2.72
r_{tt}	.544	.696	.203	.418	.480	.295	.525	.575
$P.E._{r=0}$	.1377	.0901	.1109	.0723	.1041	.0754	.1066	.1636

As a further check upon the reality of differences of reading comprehension between fast and slow readers, 50 cases in the faster-reading range were re-scored upon a basis of the questions attempted by one of the slower reading groups. The faster readers, all having attempted 29 questions (mean reading speed of about 235 W.P.M.), were compared over the same range of questions as the slower group, having attempted 22 questions (mean reading speed about 173 W.P. M.).

Presumably, if there are no real differences in reading comprehension between the fast and the slow reader, a comparison over the same range of questions would reveal that the narrowing of the range of questions over which the faster reading individuals could differ would correspondingly reduce our index of individual differences. The reverse of this was obtained. The Lexis ratio of the 50 case samples of faster readers was 1.37 compared to 1.18 for the slow readers. Reducing the range of response of the fast readers, instead of reducing the opportunity for appearance of individual differences, seemed to have actually reduced the opportunity for chance differences, while leaving the factor of individual differences in full operation.

The notion that variance of test scores is due to differences among the individuals is apparently another way of referring to the reliability of the test. Therefore we might expect the Lexis ratio to occur in some of the formulae for the estimation of test reliability. This occurrence seems to take place in Case IV of the Kuder-Richardson* formula for estimating test reliability.

Case IV of the Kuder-Richardson formula is:

$$r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma^2{}_t - n\overline{pq}}{\sigma^2{}_t}, \tag{5}$$

where n = number of items in the test,

$\bar{p}$ = mean proportion of items right,

σ_t = standard deviation of test scores (items right).

Comparing this notation with that used in the previous discussion, it may be noted that:

$n \equiv s$;

$\bar{p} \equiv p$ (proportion).

* Kuder, G. F., and Richardson, M. W. The theory of the estimation of test reliability, *Psychometrika*, 1937, 2, 151-160.

Since n may be considered a constant, the standard deviation of scores expressed as proportions of items attempted is

$$\frac{\sigma_t}{s}.$$

For the Kuder-Richardson σ^2_t, we may write $s^2(\sigma')^2$ and for $n\bar{p}\bar{q}$, the quantity $s^2(\sigma'_B)^2$ may be substituted. Then the Kuder-Richardson formula may be rewritten

$$r_{tt} = \frac{s}{s-1} \cdot \frac{s^2(\sigma)^2 - s^2(\sigma'_B)^2}{s^2(\sigma')^2}. \tag{6}$$

This reduces to

$$r_{tt} = \frac{s}{s-1}\left(1 - \frac{1}{L^2}\right). \tag{7}$$

As the Lexis ratio, L, increases, it will be seen that the value of $\frac{1}{L^2}$ will decrease toward zero as a limit and that the reliability coefficient will approach unity. In other words, as the distribution departs more and more strongly from the theoretically expected (chance) distribution, indicating a greater variance due to individual differences, the estimate of the reliability of the test is correspondingly increased. Should the ratio, L, approximate unity, the value of $\frac{1}{L^2}$ will approximate unity, so that the value of the term within the parentheses will approach zero, as will the reliability coefficient. This would indicate that the distributions obtained were what might be expected upon the basis of chance alone.

If, however, the ratio, L, decreases to less than unity, as in a Poisson distribution, it will be seen that the value of $\frac{1}{L^2}$ will increase so that the value of the term within the parentheses will be negative, as will the reliability coefficient. In connection with the possibilities of occurrence of a negative reliability coefficient, the comments of Kuder and Richardson are pertinent.

> For r_{tt} to be positive, σ^2_t must exceed $n\bar{p}\bar{q}$. Now $n\bar{p}\bar{q}$ is the variance of n equally difficult items when they are uncorrelated, by the familiar binomial theorem. Hence r_{tt} is positive for any average inter-item correlation that is positive. But negative reliability is inadmissible; hence only to the extent to which test items are positively intercorrelated

will a test have reliability. It is implicit in all formulations of the reliability problem that *reliability is the characteristic of a test possessed by virtue of the positive intercorrelations of the items composing it.*

For the eight subsamples, the reliabilities computed by formula (7) are shown in Table 1, along with probable errors. However, a better basis of judging whether or not such reliabilities could reasonably be obtained from a universe where the true value is zero can be gotten from either Fisher's z transformation or the probable error of a coefficient of correlation of zero. In order to refrain from unduly complicating the demonstration, the latter function has been used.

Table 2 shows the corresponding values for r_{tt} for selected values of L. This table was constituted on the assumption that s is large, so that $s/(s-1)$ approaches unity.

TABLE 2
Values of r_{tt} for selected values of L

L	r_{tt}
1.00	.00
1.2	.31
1.4	.49
1.6	.61
1.8	.69
2.0	.75
3.0	.89
4.0	.94
5.0	.96

PSYCHOMETRIKA—VOL. 7, NO. 4
DECEMBER, 1942

DERIVATION AND APPLICATION OF A UNIT SCORING SYSTEM FOR THE STRONG VOCATIONAL INTEREST BLANK FOR WOMEN

BERTHA P. HARPER AND JACK W. DUNLAP
UNIVERSITY OF ROCHESTER

Scoring keys, based upon unit weights, were made up for fourteen occupations of the *Strong Vocational Interest Blank for Women*. The study here presented of scores obtained in using these keys, in comparison with scores obtained from Dr. Strong's keys, indicates, for 551 women at the University of Rochester, that the new, unit-weighted keys are valid for all practical purposes and make possible a great saving in scoring time.

The *Strong Vocational Interest Blank* is a clinical instrument used in universities, personnel departments, and guidance offices. However, because of the considerable time and effort involved in scoring the test, even when machine methods are available, its use is costly and is restricted to a greater degree than is desirable. In view of these factors, a simplification of the scoring technique was proposed, involving the construction of new keys with unit weights for the item responses. This procedure was carried out in an extensive study of the *Strong Vocational Interest Blank for Men* with results that quite clearly justified the use of the unit keys; the outcome of this study was reported before the American Psychological Association in 1940 and appeared in the *Journal of Consulting Psychology* (Vol. V, no. 6, 1941). The present report deals with the application of the method to the blank for women.

The test is so constructed that a score in any occupation indicates the degree to which the subject's interests agree with those of successful individuals in that field. In effect, scoring the blank for a particular occupation is merely comparing the subject's pattern of marking, or his pattern of reactions to the items, with the typical pattern of the standardization group of individuals. Therefore, for each occupational rating, the individual's pattern of marks must be checked against a set of weights specific to each occupation. Since the women's blank contains 410 items, each with three possible responses, "like," "indifferent," and "dislike," the magnitude of the scoring task is apparent.

It is when the blank is scored by hand or on an electrical test scoring machine that the unit weights here proposed are particularly applicable. Although the men's blank is now available in machine-scored form, the women's blank has not yet been so adapted. Perhaps the basic reason why it has not been so modified is a mechanical one—that the answer sheet for use in the test scoring machine accommodates only 400 items. Therefore, the scoring of the women's blanks in this study was done on Hollerith equipment, using like, indifferent, and dislike cards for each item of the test, with a view to applying the results, if favorable, to the machine-scored method.

Using the test scoring machine with Strong's original range of weights from $+4$ to -4 necessitates two insertions of each paper into the machine for one side of the sheet, once when the machine is set so as to record values of plus or minus one and once when set to record plus or minus three. Thus, by punching the scoring stencils to the appropriate combinations of these values, any weights from $+4$ to -4 can be obtained simply by the addition of the two scores from the two separate runs. Since one side of an answer sheet can accommodate only 200 of such items, papers have to be run through the machine a total of four times, twice for each side of the sheet, and four values would have to be added in order to secure the final score for the occupation.

The simplified method of scoring of this study proposes reducing all weights of $+2$, $+3$, and $+4$ to $+1$, weights of -2, -3, and -4 to -1, and leaving weights of $+1$, 0, and -1 unchanged. Throughout this report, whenever scores are referred to as "original," they have been obtained using Strong's weights ranging from $+4$ to -4, while "unit" scores have been obtained using the new simplified weights ranging from $+1$ to -1.

Machine scoring stencils can be constructed with all weights considered as unit weights. The use of these stencils would save half the machine time since only one run is necessary for each side of the answer sheet and one third of the time in addition since only one addition is required rather than three.

The major problem in this study was to determine the effect of unit weights on the scores, and secondarily, it was necessary to eliminate ten items so that the blank could be adapted for the test scoring machine. The investigation concerned itself with fourteen scales for the women's blanks, so that there were a possible 42 effective weights for each item. The number of effective weights was determined for each item and ten items having the fewest weights were selected to be eliminated on the answer sheet. The items eliminated were 125, 141, 151, 155, 218, 228, 243, 245, 257, and 367, and the number of

effective weights for these items were, respectively, 4, 4, 3, 3, 4, 5, 3, 3, 5, and 2.

Scoring the *Strong Vocational Interest Blank for Women* on the Hollerith equipment involved first "pulling" a set of 410 cards from a file of 1230 cards, according to the way the subject marked the test (for example, a "like" card for item 1, an "indifferent" card for item 2, and so on), and then running these cards through the tabulator to obtain the "original" totals for the various occupations. Weights on the new unit method for the various occupations for the women's blank were punched into the unused remainder of the cards. Thus, once a set of cards was "pulled" for an individual, both sets of totals, for the new and old sets of weights, were obtained simultaneously for the occupations studied. The ten items that were being studied for possible elimination were always kept separately so that it was possible to secure four scores for each occupation—the original score on 400 items, the original score on 410 items, and similarly unit scores for 400 and 410 items.

All of these scores were obtained for fourteen occupations, including artist, author, librarian, secretary, lawyer, physician, nurse, social worker, Y.W.C.A. secretary, teacher in general, social science teacher, mathematics-science teacher, English teacher, and masculinity-femininity.

The subjects were 551 women students at the University of Rochester, comprising the four classes in college at the time the study was conducted. For the purpose of the study this group was divided in a random manner into two sections, one of 328 individuals which was called the "experimental group," and one of 223 individuals which was called the "control group."

The underlying methodology in the experimental design of the problem is that of correlation and regression. The validity of the unit scores was first tested by obtaining the correlation coefficients between scores obtained by the old method and the new. However, the magnitude of the correlation coefficient in itself is meaningless except on purely theoretical grounds. Considerations of practical utility demand comparisons of the *results* of the two methods, that is, of the final letter grade ratings upon the basis of which advice is given. Therefore, regression equations were constructed and applied to a new set of data (which has been called the control group), thus predicting original scores from a knowledge of the unit scores of this new group. Then, if close similarity can be demonstrated between the predicted values and the actual values, it seems reasonable to place a fair amount of confidence in the accuracy of the new method.

Before attacking the main problem in this manner it was neces-

sary to determine whether or not it is feasible to eliminate the ten items mentioned. Therefore, correlations were determined between the original scores based upon 400 items and those based upon 410 items for the entire group of 551 individuals. Eight of these were .999, three were .998, and the remaining three were .997, .989, and .985 respectively. These high correlations, when considered together with the paucity of the weights for these items, seemed justification for their elimination.

After this study had been completed, correspondence with Dr. Strong revealed that he had planned to eliminate ten items. Seven of the ten were common to both sets of items eliminated, and in the case of the other three, it was a matter of choice, since there was little or no difference in the number of effective weights. In preparing the final unit keys, the following ten items were eliminated in order to agree with Dr. Strong's revision: 131, 141, 218, 228, 236, 243, 245, 257, 348, and 367. This slight shift will in no way affect the results of the investigation, and the regression equations and tables may be used with confidence.

The next step was then obtaining the correlations between the original and unit scores in what has been called the "experimental group." In order to subject the data to the most rigorous set of cross-correlations for the purposes of constructing a machine-scored edition, in all cases the original scores based on 410 items were used ver-

TABLE 1

Correlations between Original and Unit Scores for the Experimental Group, together with Regression Coefficients and Constants

$N = 328$

Occupation	Correlation	Regression Coefficient	Regression Constant
Artist	.985	1.76	16.74
Author	.986	1.64	51.90
Librarian	.977	1.11	—7.95
Secretary	.983	1.40	—2.46
Lawyer	.976	1.42	1.70
Y.W.C.A. Secretary	.977	1.52	3.64
Social Worker	.953	1.34	—1.24
Physician	.951	1.32	4.97
Nurse	.975	1.29	—9.91
English Teacher	.961	1.33	3.57
Teacher in General	.951	1.33	—2.85
Math-Science Teacher	.966	1.38	0.60
Social Science Teacher	.959	1.33	—7.19
Masculinity-femininity	.980	1.39	—5.72

sus the unit scores based on 400 items.

The second column of Table 1 contains the correlation coefficients, which range from .986 for the author scale to .951 for physician.

Using these correlation coefficients with the corresponding means and standard deviations, regression equations were constructed for the various occupational scales for the prediction of original scores from a knowledge of unit scores. In the third and fourth columns of Table 1 appear the regression coefficients and constants that were derived.

The next step was the application of the regression equations to the unit scores of the control group of 223 individuals to predict in each case what the original score would be for that occupation. An analysis was made of the accuracy of these predictions in three ways. First, correlations were computed between the predicted values and the actual scores. From the second column of Table 2 it is seen that the best prediction in terms of the magnitude of the correlation coefficient was for the artist scale, with a value of .973, while the lowest coefficient was for teacher in general with a value of .919. These

TABLE 2

Correlations Between Predicted Original Scores and Actual Scores for the Control Group, and the Analysis of Shifts in Letter Grades When Original Scores Are Predicted, together with the Number of B Ratings That Occurred in the Scoring

$N = 223$

Occupation	Correlation	No. Changes of One-Half Letter Grade	No. Changes of One Letter Grade	No. Changes Orig. B+ to Unit Score B	Number of B Scores (Unit Sc.)
Artist	.973	35	0	6	29
Author	.971	37	0	2	27
Librarian	.945	40	2	3	24
Secretary	.960	45	0	7	44
Lawyer	.946	53	1	10	28
Y.W.C.A. Secretary	.930	56	0	7	17
Social Worker	.936	64	1	7	18
Physician	.932	49	2	3	15
Nurse	.947	67	2	8	37
English Teacher	.937	45	6	2	19
Teacher in General	.919	69	8	7	23
Math.-Science Teacher	.925	65	5	4	17
Social Science Teacher	.924	57	1	4	16
Masculinity-femininity	.927	21	0	1	8
	Total	703	28	71	322
	Per Cent	22.5	0.9	2.3	10.3

coefficients approach substantially the magnitude of the values obtained for the experimental group.

Next, an analysis was made of the shifts in letter grades that occurred in the predictions of original scores of the control group. Out of the 3122 scores that were studied in the control group (fourteen occupations times 223 individuals), 703 or 22.5 per cent were altered to the extent of one-half letter grade, as shown in the third column of Table 2. Changes of one letter grade occurred in only 28 or 0.9 per cent of the cases.

The majority of these shifts are not, however, of practical significance. It matters little whether one method rates an individual C and the other C+ or one B and the other B− . The important shifts are those occurring in the range where the counselor decides whether or not the score is sufficiently high that he should give favorable advice upon it. The general practice in guidance work usually is to give positive consideration to those occupations rated A or B+ and only doubtful consideration to lower scores. For example, Dr. John Darley, a noted authority in the clinical field, considers only scores of A and B+ in primary interest patterns; lower scores are grouped into secondary and tertiary patterns.

The crucial changes in scores are, therefore, those between B and B+. Of particular importance are those cases where the individual has an original score of B+, but, according to the unit scoring, would be rated only B. If no favorable advice is given on a B rating, the individual's attention is not called to the field. If, however, the true score is B and the unit score is B+, then slightly more emphasis is given to the occupation than is its due. This is not so serious as the failure of the counselor to mention the field. The critical cases are, thus, those of under-prediction, where the new method ranks an individual only at B when he should have been rated B+. The fifth column of Table 2 indicates that only 71 times in 3122, or two per cent of the time would the counselor have failed to mention an occupational area in using the unit method. If, however, it seems essential that as low a percentage of error as approximately one in fifty be eliminated, the alternative remains of rescoring with the original scales the papers rated B by the new method. As shown in the last column of Table 2, about ten per cent of the papers would, in that case, have to be rescored. Very frequently the counselor will utilize the individual's *pattern* of scores; the pattern of scores obtained using the unit keys is the same as that obtained with the original.

As a further verification of the method, an independent check study was carried out for the class of women entering in the fall of

1940. Both original and unit scores were obtained for 132 individuals, and the correlations between them were found to range from .949 to .991. These coefficients, as well as the means and standard deviations, were found to be of about the same magnitude as those obtained in the major study, as is shown in Table 3. Shifts of one-half letter grade occurred in 20.9 per cent of the cases (387 out of 1848 cases), as compared with 22.5 per cent in the original study, and of one letter grade in 1.2 per cent of the cases (22 out of 1848 cases), as compared with 0.9 per cent previously. Original scores of B+ would have been changed to B by the unit method in only 1.8 per cent of the cases; in other words, in this new group, advice would have been altered in one case in 56, as compared with one in 43 before. The same proportion of B scores occurred as did in the major group, 10.3 per cent. This check study seemed in every respect to verify the results obtained in the major study.

The results of the study of the use of unit-weighted scoring keys for the *Strong Vocational Interest Blank for Women* are in close agreement with those shown earlier for the blank for men. In view of the high correlations between original and unit scores, the considerable accuracy of prediction of original scores from unit scores of independent groups, and the few instances where the use of unit scores would have altered advice given to subjects, the unit-weighted method is recommended, as a means of reducing scoring time and costs and correspondingly extending the use of the test in general guidance work.

TABLE 3

Comparison of Results from Check Study with Those from Original Study

	Original Study	Check Study
Range of Correlations, Original vs. Unit Scores	.951 – .986	.949 – .991
Per Cent of Changes of One-Half Letter Grade	22.5	20.9
Per Cent of Changes of One Letter Grade	0.9	1.2
Per Cent of Changes Original Score B+ to Unit Score B	2.0	1.8
Per Cent of B Scores (Unit scoring)	10.3	10.3

[illegible]

W. G. Emmett. *An Inquiry Into the Prediction of Secondary-School Success.* University of London Press, 1942. Pp. 58.

A REVIEW

There have been many efforts to determine the value of tests in predicting academic or vocational success. The results of most of these studies are dubious because of the operation of selection. The correlations of the initial tests and the criteria of success and the regression coefficients used in prediction are smaller than they would be if selection had not operated. The reason for this state of affairs is that criterion measures are not usually obtainable for all members of the population taking the initial tests. The research reported by Emmett is unique in that techniques are applied which yield estimates of the regression coefficients, their standard errors, and the correlations which would be obtained had selection not operated. Use was made of Aitken's matrix formulation of Karl Pearson's selection equations. The procedure is illustrated for three independent variables—intelligence quotient, English score, and arithmetic score. Appendix III of the monograph shows how Aitken's method of pivotal condensation can be used when extending the procedure to more than three independent variables. This scholarly little monograph should be required reading for all persons engaged in attempts to use regression equations in the prediction of academic or vocational success.

MAX D. ENGELHART.

PSYCHOMETRIKA—VOL. 7, NO. 4
DECEMBER, 1942

INDEX FOR VOLUME 7

AUTHORS

Bloom, Benjamin S. (with Ardie Lubin), "Use of the Test Scoring Machine and the Graphic Item Counter for Statistical Work," 233-241.

Deemer, Walter L., "A Method of Estimating Accuracy of Test Scoring," 65-73.

DuBois, Philip H., "Note on the Computation of Biserial r in Item Validation," 143-146.

Dunlap, Jack W., "The Psychometric Society—Roots and Powers," 1-8.

Dunlap, Jack W. (with Bertha P. Harper), "Derivation and Application of a Unit Scoring System for the Strong Vocational Interest Blank for Women," 289-295.

Edgerton, Harold A. (with Kenneth F. Thomson), "Test Scores Examined with the Lexis Ratio," 281-288.

Engelhart, Max D., "Unique Types of Achievement Test Exercises," 103-115.

Engelhart, Max D., A Review of "An Inquiry into the Prediction of Secondary-School Success," by W. G. Emmett, 297.

Ferguson, George A., "Item Selection by the Constant Process," 19-29.

Ferguson, Leonard W. (with Warren R. Lawrence), "An Appraisal of the Validity of the Factor Loadings Employed in the Construction of the Primary Social Attitude Scales," 135-138.

Grossnickle, Louise T., "The Scaling of Test Scores by the Method of Paired Comparisons," 43-64.

Guilford, J. P. (with Thoburn C. Lyons), "On Determining the Reliability and Significance of a Tetrachoric Coefficient of Correlation," 243-249.

Gulliksen, Harold, "An Analysis of Learning Data Which Distinguishes Between Initial Preference and Learning Ability," 171-194.

Harper, Bertha P. (with Jack W. Dunlap), "Derivation and Application of a Unit Scoring System for the Strong Vocational Interest Blank for Women," 289-295.

Heese, K. W., "A General Factor in Improvement with Practice," 213-223.

Holzinger, Karl J., "Why Do People Factor?" 147-156.

Jackson, Robert W. B., "Note on the Relationship Between Internal Consistency and Test-Retest Estimates of the Reliability of a Test," 157-164.

Karlin, J. E., "A Factorial Study of Auditory Function," 251-279.

Katzoff, E. T., "The Measurement of Conformity," 31-42.

Kelley, T. L., "The Reliability Coefficient," 75-83.

Lawrence, Warren R. (with Leonard W. Ferguson), "An Appraisal of the Validity of the Factor Loadings Employed in the Construction of the Primary Social Attitude Scales," 135-138.

Libby, J. E. P., "Response Relay," 139-141.

Lubin, Ardie (with Benjamin S. Bloom), "Use of the Test Scoring Machine and the Graphic Item Counter for Statistical Work," 233-241.

Lyons, Thoburn C. (with J. P. Guilford), "On Determining the Reliability and Significance of a Tetrachoric Coefficient of Correlation," 243-249.

McCloy, C. H., " 'Blocks Test' of Multiple Response," 165-169.

McNemar, Quinn, "On the Number of Factors," 9-18.

Rashevsky, N., "Contributions to the Mathematical Theory of Human Relations: V.," 117-134.

Rashevsky, N., "Further Studies on the Mathematical Theory of Interaction of Individuals in a Social Group," 225-232.

Thomson, Kenneth F. (with Harold A. Edgerton), "Test Scores Examined with the Lexis Ratio," 281-288.

Thorndike, Robert L., "Regression Fallacies in the Matched Groups Experiment," 85-102.

Tsao, Fei, "Tests of Statistical Hypotheses in the Case of Unequal or Disproportionate Numbers of Observations in the Subclasses," 195-212.